Incorporating Covariates in Skewed Functional Data Models

Meng Li, Ana-Maria Staicu* and Howard D. Bondell

Department of Statistics, North Carolina State University
ana-maria.staicu@ncsu.edu

November 5, 2014

Abstract

We introduce a class of covariate-adjusted skewed functional models (cSFM) designed for functional data exhibiting location-dependent marginal distributions. We propose a semi-parametric copula model for the pointwise marginal distributions, which are allowed to depend on covariates, and the functional dependence, which is assumed covariate invariant. The proposed cSFM framework provides a unifying platform for pointwise quantile estimation and trajectory prediction. We consider a computationally feasible procedure that handles densely as well as sparsely observed functional data. The methods are examined numerically using simulations and is applied to a new tractography study of multiple sclerosis. Furthermore, the methodology is implemented in the R package cSFM, which is publicly available on CRAN.

Keywords: Covariate modeling, diffusion tensor imaging, functional principal component analysis; Gaussian copula; quantile estimation; skewed func

*To whom correspondence should be addressed.
1 Introduction

Development in technology and computation has facilitated the recording of repeated measures with high frequency, allowing the data to be viewed as a function. Functional data analysis (FDA) has developed rapidly due to the increasing applications in many areas such as medical fields, environmental science, image analysis, and traffic modeling (Ramsay and Silverman, 2002, 2005; Ferraty and Vieu, 2006; Serban and others, 2013; Gertheiss and others, 2013, to name a few). Most approaches in FDA are based on the first two order moments - mean and covariance functions. An example is the functional principal component analysis (FPCA), which has become a standard tool to achieve dimension reduction in FDA. Limited investigation has been done to incorporate covariates in the framework of FPCA (Cardot, 2007; Jiang and Wang, 2010). However, there are important cases where both higher order moments and covariate information are of interest. For example, in our brain tractography application it is of interest to study how the parallel diffusivity recorded at many locations along the main white matter tract of the brain - corpus callosum - varies with the tract location, while accounting for additional covariate information provided by the parallel diffusivity mean summary along a neighboring tract - left corticospinal tract. Furthermore we want to study how this dependence is different for multiple sclerosis (MS) diseased and healthy subjects; see Figure 1 for visual display. At a quick look it appears that the skewness of the parallel diffusivity along the corpus callosum varies across the tract location, and this dependence is different for MS and healthy subjects. Thus, accounting for higher-order moments of the pointwise distribution, as well as for the additional available covariate in the modeling of the parallel diffusivity profiles has the potential to provide a more complete description of the observed data, which could lead to generating new scientific hypotheses. In this paper, we introduce a novel semiparametric approach to model functional data in the presence of skewed distributions, while accounting for additional covariates, and propose a practically feasible estimation algorithm.

The approach we propose is fundamentally different from the standard FPCA because: 1) it
targets the entire distribution of the functional data; 2) it uses copula to account for functional
dependence; 3) it allows for non-symmetric location-varying distribution of the data, and 4) it
accommodates additional covariate information in a parsimonious manner. Staicu and others
(2012) introduced a copula-based approach and proposed models for functional data exhibiting
skewness that vary with temporal location. Copula approaches (Sklar, 1959) have been under
intense methodological development and allow to separate the modeling into the pointwise distri-
butions and the dependence structure. While the pointwise distributions capture the higher order
moments, the copula describes the dependence. One advantage of this approach is that it pro-
vides an unifying platform for both pointwise quantile estimation and prediction of trajectories
by incorporating functional dependence. Our paper considers this direction; the proposed model
framework extends Staicu and others (2012) to incorporate additional covariates in a parsimonious
way. Recently, Soiaporn and others (2013) proposed an extension of the copula-based approach to
model jointly multiple functional responses.

When covariate information is available, Cardot (2007) proposed conditional FPCA, by condi-
tioning on the covariate, and discussed estimation of the conditional mean and covariances through
nonparametric kernel estimators. Jiang and Wang (2010) provided a covariate-adjusted FPCA,
by assuming covariates effect either in the mean and/or covariance functions. Both approaches
rely on the first two moments to either estimate the mean and covariance functions or reconstruct
the trajectories, and do not account for functional data that are pointwise skewed. The methods
introduced in Staicu and others (2012), which account for such features are not directly applicable,
as they require multiple responses to be observed for each time point and covariate value, which is
almost never the case when the covariate is continuous - the case of interest in this paper.

We consider flexible parametric families to describe the pointwise distributions and assume that
the corresponding mean, variance, and additional parameters describing the shape of the distri-
bution vary smoothly with both the time and the additional covariate. To ensure a parsimonious
model, we assume that the dependence - as accounted for, via copula - is covariate-invariant. Nevertheless, our modeling strategy allows for the (Pearson) covariance of the functional responses to vary with the additional covariate (Cardot, 2007; Jiang and Wang, 2010). Our estimation approach is based on modeling the smooth parameter functions using bi-variate regression splines and employing computationally feasible simultaneous optimization for the unknown basis coefficients; the knots selection is carried using information criterion. Once the parameter functions are estimated, pointwise quantile estimates can be obtained directly, at no additional cost, for any level of quantile. Prediction of the trajectories is based on the calibration of the process’ dependence, which allows to borrow strength from the existing FPCA techniques. The methods are implemented in the R package cSFM, which is publicly available.

The remaining of this paper is organized as follows. Section 2 introduces the proposed covariate-adjusted skewed functional model. Section 3 discusses an estimation technique using regression splines, as well as the model selection. Section 4 describes the prediction procedure. The finite sample performance of the proposed method is evaluated through simulation studies in Section 5. Section 6 demonstrates an application of the proposed method to the tractography data.

2 Modeling Methodology

Let \( Y_{ij} \) be the response for subject \( i \) measured at time point \( t_{ij} \in T \) for \( j = 1, \ldots, m_i \) and let \( X_i \) be a \( q \)-dimensional vector of additional covariate information associated with the subject, for \( i = 1, \ldots, n \). With slight abuse of notation let \( Y_i(\cdot) = Y_i(\cdot, X_i) \) be the underlying function associated with the \( i \)th subject such that \( Y_{ij} = Y_i(t_{ij}) \); here the notation \( Y_i(\cdot, X_i) \) is used to emphasize the dependence of the response on the covariate \( X_i \). It is assumed that \( Y_i(\cdot, x) \) is a square integrable function for all \( x \) and that \( T \) is closed interval. We introduce a semiparametric modeling framework to describe covariate-dependent functional data as

\[
Y_i(t, X_i) = \mu(t, X_i) + \sigma(t, X_i)G^{-1}\{U_i(t); \alpha(t, X_i)\} \quad (i = 1, 2, \ldots, n),
\]  

(1)
where \( \mu(t, x) \) and \( \sigma(t, x) \) denote the mean and standard deviation of \( Y_i(t; x) \), \( G(\cdot, \alpha) \) is a cumulative distribution function (CDF) with zero-mean, unit-variance and which is parameterized by parameter, scalar or vector, \( \alpha \), and \( G^{-1}(\cdot, \alpha) \) is its inverse function. Because \( \alpha \) does not affect the mean nor the variance of the distribution, but instead is related to higher order moments - such as skewness and kurtosis - it is referred to as the “shape parameter”. For example, \( G(\cdot, \alpha) \) could be the CDF of a standard skewed normal distribution (Azzalini, 1985) parameterized by a scalar parameter or the skew-\( t \) distribution (Azzalini, 2013), parameterized by a two-dimensional shape parameter. Model (1) assumes that the shape parameter \( \alpha \) varies over \( t \) and in addition depends on the covariate \( X_i \). For each \( t \in T \), the random variable \( U_i(t) \) is assumed to follow \( \text{Uniform}[0, 1] \). The dependence of \( Y_i(\cdot, X_i) \) is modeled implicitly through the correlated latent process \( U_i(\cdot) \); thus (1) assumes that the covariate influences the pointwise distributions, but not the dependence, as captured by \( U_i(t) \). The dependence will be modeled using copulas, such as Gaussian copula (Ruppert, 2010), which allow modularity, by separating pointwise distributions and the mutual stochastic dependence. Although model (1) is parsimonious, remark that it allows the covariate \( X_i \) to influence both the mean and the Pearson covariance of the process \( Y_i(\cdot, X_i) \).

The class described by model (1) extends the class of models proposed by Staicu and others (2012) for skewed functional data, to account for additional covariate information; thus we call it covariate-adjusted skewed functional model (cSFM). For discrete covariate \( X_i \), the approach of Staicu and others (2012) can still be used, provided that the number of realizations \( Y_i(\cdot, X_i) \) for each covariate value increases with the sample size. Nevertheless, when the covariate \( X_i \) is continuous the methods discussed in Staicu and others (2012) are not applicable and new methodology is required. The extension raises many challenges and is far from straightforward. For example, accounting for covariates as proposed in model (1) leads to an increased dimensionality of the problem, and thus demands for computationally feasible algorithms. Models for functional data that accommodate covariate effects have been investigated recently. For example, Jiang and Wang
(2011) proposed single index models to summarize the covariate effect in the mean function, and consider regular FPCA to model the residuals; Jiang and Wang (2010) introduced a modification of the existing FPCA to account for additional covariates by incorporating covariate effects in the mean and/or covariance functions. However, the current approaches are limited in that the complete pointwise distribution is implicitly assumed to be characterized by its first two moments. However, even if the covariate only affects these two moments, the higher order moments should be taken into account during estimation.

A close inspection of the model (1) shows that if $\mu(t, x), \sigma(t, x)$ and $\alpha(t, x)$ are known, then $U_i(t) = G\left\{ Y_i(t, X_i) - \mu(t, X_i) \right\}/\sigma(t, X_i); \alpha(t, X_i) \right\}$ is the latent process obtained by a transformation of the observed functional data $Y_i(t, X_i)$. Thus, model (1) serves a dual purpose: 1) to generalize standard approaches to functional data, by providing a distributional framework; and 2) to provide a class of transformations for functional data that maps correlated functional processes with complex pointwise characteristics to correlated functional processes with simple pointwise characteristics. Also (1) provides a unifying platform for 1) modeling the pointwise quantile functions as $Q_p(t, x) = \mu(t, x) + \sigma(t, x)G^{-1}\{p; \alpha(t, x)\}$, for $p \in (0, 1)$ and 2) reconstruction of individual profiles, using standard FPCA techniques to recover the latent process $U_i(\cdot)$.

3 Estimation Procedure

We estimate the population level functions $\mu(\cdot, \cdot), \sigma(\cdot, \cdot)$ and $\alpha(\cdot, \cdot)$ using regression splines, based on tensor product B-splines; approaches based on the other bases choices may be used following the same scheme. Since $U_i(t)$ is uniformly distributed for each $t$, it follows that the joint distribution of $\{U_i(t_{i1}), \ldots, U_i(t_{im})\}$ is, by definition, a copula; parametric copulas, such as Gaussian copulas, that are parameterized by correlation functions are considered. The estimation methodology contains two steps: 1) estimation of the parameters jointly by a pseudo-likelihood approach, and 2) nonparametric estimation of the latent copula correlation.
3.1 Population level function estimation

The methodology is demonstrated via a univariate continuous covariate, while extensions to discrete and categorical covariates, and the multiple covariate case follow naturally. Let \( \{B_{T,l}(t) : 1 \leq l \leq K_t\} \) and \( \{B_{X,k}(x) : 1 \leq k \leq K_x\} \) be univariate B-spline bases of degrees \( s_t \) and \( s_x \) on \( T \) and \( X \) respectively, with \( q_t \) and \( q_x \) interior knots placed at equally spaced quantiles (de Boor, 2001; Ruppert and others, 2003); here \( K_t = s_t + q_s + 1 \) and \( K_x = s_x + q_x + 1 \). We use a bivariate spline model representation based on tensor product B-spline bases (Durrett, 2005) for both the mean and log variance functions:

\[
\mu(t, x) = \sum_{l=1}^{K_t} \sum_{k=1}^{K_x} \beta_{\mu, l} B_{T,l}(t) B_{X,k}(x), \quad \log\{\sigma^2(t, x)\} = \sum_{l=1}^{K_t} \sum_{k=1}^{K_x} \beta_{\sigma, l} B_{T,l}(t) B_{X,k}(x),
\]

where the numbers of univariate bases functions, \( K_t \) and \( K_x \), for the mean and variance functions are assumed equal, for simplicity. Often the shape parameter function requires careful handling. In our experience modeling an appropriate one-to-one transformation improves the numerical stability (see also Staicu and others, 2012). Let \( h(\cdot) : R \rightarrow R \) be such a transformation;

\[
h\{\alpha(t, x)\} = \sum_{l=1}^{K_t} \sum_{k=1}^{K_x} \beta_{\alpha, l} B_{T,l}(t) B_{X,k}(x);
\]

for convenience, assume the same number of univariate basis functions, \( K_t \) and \( K_x \), as for the mean/variance functions. Different bases are used in the simulation and allowed in the developed \texttt{R} package \texttt{cSFM}. If \( G(\cdot, \alpha) \) is the centered/scaled skewed normal distribution, we work with the centered parameterization, based on skewness (Arellano-Valle and Azzalini, 2008). Let \( h(\alpha) = \Phi^{-1}\{(M^{-1}\gamma_\alpha + 1)/2\} \), for \( \Phi^{-1}(\cdot) \) is the inverse CDF of the standard normal, \( M = \sqrt{2(4 - \pi)(\pi - 2)^{-3/2}} \), and \( \gamma_\alpha = \sqrt{2(4 - \pi)}\delta^3(\pi - 2\delta^2)^{-3/2} \) for \( \delta = \alpha/\sqrt{1 + \alpha^2} \). Function \( h(\cdot) \) reparameterizes the shape parameter using the skewness \( \gamma_\alpha \), which is bounded by \( M \), and maps it to the real line.

Let \( \beta_\mu \) be the \( K_x K_t \)-dimensional column vector obtained by stacking \( \beta_{\mu, kl} \) first over \( k \) and then over \( l \), and similarly define the vectors \( \beta_\sigma \) and \( \beta_\alpha \). As well define \( B_{\mu, ij} \) as the \( K_x K_t \)-dimensional column vector with elements \( B_{T,l}(t_{ij})B_{X,k}(X_i) \), so we have \( \mu(t_{ij}, X_i) = B_{\mu, ij}^T \beta_\mu \). Likewise define the
$K_xK_t$ dimensional vectors $B_{\sigma,ij}$ and $B_{\alpha,ij}$ such that $\log\{\sigma^2(t_{ij}, X_i)\} = B_{\sigma,ij}^T\beta_{\sigma}$ and $h\{\alpha(t_{ij}, X_i)\} = B_{\alpha,ij}^T\beta_{\alpha}$ respectively. It follows that observation $Y_{ij}$ has mean equal to $\mu_{ij} = B_{\sigma,ij}^T\beta_{\mu}$, variance equal to $\sigma_{ij}^2 = \exp(B_{\sigma,ij}^T\beta_{\sigma})$ and furthermore that $(Y_{ij} - \mu_{ij})/\sigma_{ij}$ has distribution with the CDF specified by $G(\cdot; \alpha_{ij})$ where $\alpha_{ij} = h^{-1}(B_{\alpha,ij}^T\beta_{\alpha})$, and $h^{-1}(\cdot)$ is the inverse function of $h(\cdot)$. Conditional on the bivariate basis functions, the parameters $\beta_{\mu}$, $\beta_{\sigma}$ and $\beta_{\alpha}$ are estimated using maximum likelihood estimation and a working independence assumption. Specifically, if $\beta = (\beta_{\mu}^T, \beta_{\sigma}^T, \beta_{\alpha}^T)^T$ is the overall parameter vector, and denote by $p$ its length, then $\hat{\beta} = \arg \max_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \sum_{j=1}^{m_i} \ell_{ij}(\beta; Y_{ij})$, where $\ell_{ij}(\beta; Y_{ij}) = \log g\{(Y_{ij} - B_{\mu,ij}^T\beta_{\mu}) \exp(-B_{\sigma,ij}^T\beta_{\sigma}/2); h^{-1}(B_{\alpha,ij}^T\beta_{\alpha})\} - B_{\sigma,ij}^T\beta_{\sigma}/2$ is the log-likelihood function of $Y_{ij}$ and $g(\cdot; \alpha)$ is the probability density function corresponding to the CDF $G(\cdot; \alpha)$. The estimates of the parameter functions $\hat{\mu}(\cdot, \cdot)$, $\hat{\sigma}(\cdot, \cdot)$ and $\hat{\alpha}(\cdot, \cdot)$ are obtained by substituting $\beta_{\mu}$, $\beta_{\sigma}$, and $\beta_{\alpha}$ from (2) and (3) with the corresponding counterparts from $\hat{\beta}$.

Quasi-Newton algorithm can be used for the likelihood optimization via function \texttt{optim} in the R package \texttt{stats}. However, a direct application of this function is unstable due to the large dimension of the parameters. In our developed R package and simulation experiments, the parameter estimation is carried with \texttt{optim} by providing closed form expressions for the gradient, leading to stable computation and faster iteration convergence rates.

The function $\ell_{\text{ind}}(\beta; Y) = \sum_{i=1}^n \sum_{j=1}^{m_i} \ell_{ij}(\beta; Y_{ij})$ is a misspecified likelihood, based on an independence model; $Y$ is the vector of $Y_{ij}$’s. Using Sklar’s theorem (Sklar, 1959), the full likelihood of the observed data $Y$ can be written as $\ell_{\text{full}}(\beta, \Omega; Y) = \ell_{\text{ind}}(\beta; Y) + \ell_C(\beta, \Omega; Y)$, where $\ell_C(\cdot, \cdot; Y)$ is the log-likelihood function of the copula and is fully specified by the parameters $\beta$ and $\Omega$. For the Gaussian copula, the parameter $\Omega$ is the correlation of an appropriate latent random variable. Direct estimation of $\beta$ and $\Omega$ using the full likelihood involves optimization over a large dimensional space, and thus would be computationally challenging or unfeasible. Our proposed estimation - known in the literature by “pseudo likelihood” (Ruppert, 2010) - uses two steps. First, the population-level parameters are estimated using $\ell_{\text{ind}}(\beta; Y)$, and then the copula parameter is
estimated as the correlation of an appropriate latent variable, using a method of moment approach.

Once the population level functions are estimated, the cSFM approach allows the estimation of the pointwise quantile for any quantile levels: the estimated pointwise quantile function of level \( \tau \) is obtained by 
\[
\hat{Q}_\tau(t, x) = \hat{\mu}(t, x) + \hat{\sigma}(t, x) G^{-1}\{\tau; \hat{\alpha}(t, x)\}.
\]

3.2 Copula calibration

Here we discuss estimation of the copula parameter \( \Omega \), when a Gaussian copula is assumed to describe the dependence. We begin with providing insights into the latent variable whose correlation is given by the copula parameter. Specifically, let 
\[
Z_i(t_{ij}) = \Phi^{-1}\{U_i(t_{ij})\},
\]
where recall 
\[
U_i(t) = G\{Y_i(t, X_i) - \mu(t, X_i)\}/\sigma(t, X_i); \alpha(t, X_i)\}
\]
is the covariate-free underlying process. It follows that \( Z_i(t_{ij}) \) has standard normal distribution, for every \( t_{ij} \). Furthermore, let \( K(\cdot, \cdot) \) be the covariance function of the induced Gaussian process \( \Phi^{-1}\{U_i(t)\} \) defined by 
\[
K(t, t') = \text{cov}[\Phi^{-1}\{U_i(t)\}, \Phi^{-1}\{U_i(t')\}];
\]
notice that the covariance and correlation function coincide for the induced latent process. Thus, the copula parameter, \( \Omega \), consists of all correlation coefficients 
\[
cor\{Z_i(t_{ij}), Z_i(t_{ij}')\} = K(t_{ij}, t_{ij}'), \text{ for all } i, j.
\]
Due to the large dimensionality of this parameter, likelihood-based approaches would be unfeasible. Instead, moment based methods are a common alternative in practice (Ruppert, 2010). Our approach relies on the assumption that the covariance function \( K(t, t') \) is smooth for \( t \neq t' \), which is commonly made in the FDA literature (Yao and others, 2005; Ramsay and Silverman, 2005). For given estimates of the population level functions, the copula parameter can be estimated directly using either 1) the Pearson correlation of the induced latent process or 2) the nonparametric Kendall’s tau; we detail each method in part next.

**Pearson correlation.** Let \( \hat{\mu}_{ij}, \hat{\sigma}_{ij}, \) and \( \hat{\alpha}_{ij} \) be the estimates of the mean, standard deviation and shape based on the fixed bases and the parameter estimate \( \hat{\beta} \). The induced latent Gaussian process evaluated at time \( t_{ij} \) can be approximated by 
\[
\tilde{Z}_i(t_{ij}) = \Phi^{-1}\{(Y_{ij} - \hat{\mu}_{ij})/\hat{\sigma}_{ij}; \hat{\alpha}_{ij}\}.
\]
To account for all sources of errors, it is assumed that \( \tilde{Z}_i(t_{ij}) \) is a realization of a Gaussian latent process at time \( t_{ij} \) which is contaminated with measurement error. Standard FPCA techniques can
be used to estimate the correlation of the latent process, via a reduced rank approximation of the covariance and bivariate smoothing techniques. Let \( \hat{\lambda}_k \) and \( \hat{\phi}_k(t) \) be the estimated \( k \)th eigenvalue and eigenfunction and \( \hat{\sigma}_e^2 \) be the estimated noise variance. The underlying covariance \( K(\cdot, \cdot) \) is estimated by \( \hat{K}(t, t') = \sum_{k=1}^{L} \hat{\lambda}_k \hat{\phi}_k(t) \hat{\phi}_k(t') + \hat{\sigma}_e^2 1(t = t') \), where \( L \) is a finite truncation determined by the percentage of explained variance criterion; see Staicu and others (2012) for more details.

**Kendall’s tau.** The correlation function \( K(\cdot, \cdot) \) can also be estimated by the Kendall’s tau correlation of the latent process. Using (5.32) of McNeil and others (2010), let \( \tilde{\rho}_\tau(t_j, t'_j) \) be the sample Kendall’s tau between \( \{U_i(t_{ij}) : t_{ij} = t_j\} \) and \( \{U_i(t_{ij}) : t_{ij} = t'_{ij}\} \); \( \tilde{G}(t_j, t_{j'}) = \sin\{\pi \tilde{\rho}_\tau(t_j, t'_{j})/2\} \) can be viewed as an undersmooth estimator of the Pearson correlation of the latent Gaussian process contaminated with error. Reduced rank approximation is again used to ensure positive definiteness of the estimated correlation function. The use of Kendall’s tau is more appealing because of the invariance property of Kendall’s tau to increasing functions.

When small values of the response at points \( t_0 \) are followed by small values at the adjacent points, then a \( t \)-copula captures the model dependence better. Our methodology can be extended to accommodate a \( t \)-copula. The estimation of the copula parameters follows roughly the same direction as Staicu and others (2012): the main difference is that both the Kendall’s tau correlation and the degrees of freedom parameter will be estimated from the underlying process \( \tilde{U}_i(t) \), where \( \tilde{U}_i(t_{ij}) = G\{(Y_{ij} - \hat{\mu}_{ij})/\hat{\sigma}_{ij}; \hat{\alpha}_{ij}\} \) as opposed to the response \( Y \).

### 3.3 Selection of the number of knots

Our estimation algorithm is based on specified bases functions, as described by preset degrees and number of knots that are placed at equally spaced quantiles. Typically, the degree/order of the basis functions is chosen to describe certain characteristics of the parameter functions - such as the number of continuous derivatives. The number of knots is a tuning parameter and controls the bias-variance trade-off. We propose to select the number of knots using the Akaike information criterion (AIC) (Akaike, 1970), by exploiting the semi-parametric modeling framework. In general,
AIC = $-2\ell_{full}(\hat{\beta}, \hat{\Omega}; Y) + 2p$, and the optimal number of knots is selected by minimizing AIC. AIC does not depend on the sample size - a property which makes it more appealing when comparing with its competitors Bayesian information criterion and cross-validation (CV) methods. Also AIC is shown to be asymptotically optimal in terms of choosing the best approximating model when the underlying true model is infinite-dimensional (Shibata, 1981; Li, 1987).

4 Prediction of Trajectories

Reconstructing trajectories is highly important in functional data analysis. Benefiting from the copula based modeling in (1), we can reconstruct trajectories by borrowing strength from the standard FPCA techniques. Specifically, let $Y^{new}(\cdot, X_0)$ be a new response observed incompletely at few values $\{t_1, \ldots, t_{m_{new}}\}$, and corresponding to the covariate value $X_0$. To reconstruct the trajectory $Y^{new}(\cdot, X_0)$, the key step is in the prediction of the latent trajectory $\hat{Z}^{new}(\cdot)$. Once such $\hat{Z}^{new}(\cdot)$ is available, let $\hat{U}^{new}(\cdot) = \Phi\{\hat{Z}^{new}(\cdot)\}$ and calculate the predicted trajectory $\hat{Y}^{new}(t, X_0)$ using model (1) with the estimated population level functions $\hat{\mu}(t, X_0), \hat{\sigma}(t, X_0), \hat{\alpha}(t, X_0)$ and the predicted latent trajectory $U^{new}(t)$ in place of $\mu(t, X), \sigma(t, X), \alpha(t, X)$, and $U(t)$ respectively; $\hat{\mu}, \hat{\sigma}$ and $\hat{\alpha}$ are determined as in Section 3.1.

Next, we discuss prediction of the latent trajectory $\hat{Z}^{new}(\cdot)$. For any $t \in \{t_1, \ldots, t_{m_{new}}\}$ application of the previous formula gives $\hat{Z}^{new}(t) = \Phi^{-1}(G\{Y^{new}(t, X_0) - \hat{\mu}(t, X_0)\}/\hat{\sigma}(t, X_0); \hat{\alpha}(t, X_0))$. As discussed in Section 3.2, our methodology assumes that the latent random curves $Z_i(\cdot)$ are independent and identical realizations of a process with smooth covariance function that are contaminated with measurement error. It can be decomposed using the Karhunen-Loève expansion as $Z_i(t) = \sum_{k \geq 1} \xi_{ik} \phi_k(t) + \epsilon_i(t)$, where $\xi_{ik}$ are zero-mean random variables, with covariance equal to $\lambda_k$, and uncorrelated over $k$, and $\epsilon_i(t)$ is white noise process with covariance $\text{cov}\{\epsilon(t), \epsilon_i(t')\} = \sigma^2(t = t')$; here $\lambda_1 > \lambda_2 > \ldots \geq 0$. Using the observed data $\{(Y_{ij}, t_{ij}) : j, X_i\}_i$, and thus $\{\tilde{Z}_i(t_{ij})\}_i$, and furthermore let $\tilde{Z}^{new}(t) : t \in \{t_1, \ldots, t_{m_{new}}\}$ correspond to the incompletely observed $Y^{new}(t, X_0) : t \in \{t_1, \ldots, t_{m_{new}}\}$. We predict $\tilde{Z}^{new}(\cdot)$ by $\tilde{Z}^{new}(t) = \sum_{k=1}^L \hat{\phi}_k(t) \tilde{\xi}^{new}_k$, where
\( \hat{\phi}_k(t) \) are the estimated eigenfunctions corresponding to the sample of discretely and noisy measured profiles \( \{ \tilde{Z}_i(t_{ij}) \}_i \) and \( L \) is the reduced rank truncation used in the estimation of the copula correlation, as determined in Section 3.2. Furthermore \( \hat{\xi}_k^{\text{new}} = E[\xi_k|\tilde{Z}^{\text{new}}(t_1), \ldots, \tilde{Z}^{\text{new}}(t_{m_{\text{new}}})] \) is determined using conditional expectation (Yao and others, 2005).

Intuitively, the proposed methodology is directly applicable to functional data observed on a sparse design; however additional numerical investigation is necessary in order to assess the accuracy of the estimation procedure in this setting.

5 Simulation Results

We conducted a simulation experiment to show the estimation and predictive performance of the proposed method for increasing sample sizes, and compare the results with several alternatives. Specifically, we generate data \( \{Y_{ij}, t_{ij} : 1 \leq j \leq m_i \}, X_i \) where \( Y_{ij} = Y_i(t_{ij}, X_i) \) using model (1) with model components detailed below. For each \( i \), the timepoints \( \{t_{ij} : j \} \) are 80 equispaced grid of values in \([0, 1]\), and the covariate \( X_i \) is generated from Uniform\([0, 1]\). The pointwise mean and variance functions are taken as \( \mu(x, t) = \sin(\pi x) \cos(\pi t), \sigma(x, t) = 25 \exp(-8)\phi(5x-2.5)\phi(5t-2.5), \) where \( \phi(\cdot) \) is the probability density function for a standard normal variable. We take \( G\{\cdot; \alpha(t)\} \) to be the CDF of the centered and scaled skewed normal distribution (Azzalini, 2013) with shape parameter \( \alpha(t) = 10 \sin(2\pi t) \). Furthermore, the stochastic process \( U_i \) is generated from \( U_i(t) = F_{V_i(t)}\{V_i(t)\} \) where \( V_i(t) = Z_i(t) + \epsilon_i(t) \), \( Z_i \) is zero-mean Gaussian process with covariance function \( \text{cov}\{Z_i(t), Z_i(s)\} = \sum_{l \geq 1} \lambda_l \varphi_l(t)\varphi_l(s) \), \( \epsilon_i(t) \) is independently generated as \( N(0, \sigma^2 = 0.10) \) for all \( i \) and \( t \), and \( F_{V_i(t)} \) is the CDF of \( V_i(t) \). Here \( \{\varphi_l(\cdot)\}_{l \geq 1} \) is the Fourier basis with \( \varphi_1(t) = \sqrt{2}\sin(2\pi t), \varphi_2(t) = \sqrt{2}\cos(4\pi t), \varphi_3(t) = \sqrt{2}\sin(4\pi t) \), and so on, and \( \lambda_l = (1/2)^{l-1} \) for \( l = 1, 2, 3 \) and 0 otherwise. For each scenario considered, results are based on 100 simulations.

We measure the performance of all compared methods in three ways: 1) estimation of the population level functions, \( \mu(t, x), \sigma(t, x), \) and \( \alpha(t), \) 2) prediction performance, as well as 3) estimation of pointwise quantile functions. To assess 1) and 2) we compare the cSFM, by assuming
that $G\{\cdot; \alpha(t)\}$ is in the family of skewed Normal distributions with time-varying shape parameter $\alpha(t)$, with three alternative approaches: cSFM$_0$, which is a variant of the cSFM with the shape parameter is set to 0; two-step cSFM, which is a two-step procedure that combines penalized bivariate splines for the estimation of the mean and cFSM method for the de-meaned data; and the mean covariate adjusted FPCA method (mFPCA) introduced by Jiang and Wang (2010), which assumes the model $Y_i(t,X_i) = \mu(t,X_i) + \sum_{k \geq 1} \zeta_{ik}\psi_k(t) + \varepsilon_i(t)$ using standard FPCA notations. Here bivariate penalized spline smoothing is used to model the mean function, and common FPCA techniques are used for the pseudo-residuals $\tilde{Y}_{ij} = Y_i(t_{ij},X_i) - \hat{\mu}(t_{ij},X_i)$. To assess 3), we compare the pointwise quantile functions estimated using cSFM for various quantile levels with the corresponding counterparts via cSFM$_0$, two-step cSFM. The pointwise quantile functions with mFPCA are estimated based on Gaussian assumption of the FPC scores, which in turn imply a Gaussian distribution for the response; note that mFPCA does not necessarily assume Gaussian scores when used to make prediction. Additionally, we include the pointwise quantiles estimated using the constrained B-splines nonparametric regression quantiles (COBS) method (He and Ng, 1999).

The model fitting is based on cubic regression splines to model the mean, variance and shape parameter functions combined with Gaussian copula to model the process dependence, and it employs the methodology described in Section 3. For each parameter functions, $\mu(\cdot,\cdot), \sigma(\cdot,\cdot)$ and, $\alpha(\cdot,\cdot)$, we use the same number of knots in the $x$- and $t$- directions, $K\mu, K\sigma$, and $K\alpha$ respectively; the optimal number of knots is selected via AIC using a grid search with the restriction $K\mu > K\sigma > K\alpha$ (Wang and others, 2008). For both the two-step cSFM and mFPCA, the mean function is modeled using penalized splines instead and the smoothness is controlled via two smoothing parameters; fitting is done via the R package mgcv of Wood (2011), and the smoothing parameters are selected by REML. COBS is implemented via the R package cobs (Ng and Maechler, 2011), employing quadratic splines and Schwarz-type information criterion (for the smoothing parameter).

The performance estimation of model components is assessed through the square root of inte-
grated mean squared error ($\sqrt{\text{IMSE}}$), which, for some generic bi-variate function $\theta(t, x)$ with estimator $\hat{\theta}(t, x)$ is defined as $\text{IMSE}(\theta) = \int_0^1 \int_0^1 \mathbb{E}[(\hat{\theta}(t, x) - \theta(t, x))^2] dtdx$. Additionally, we consider the Kullback-Leibler divergence (KL) to evaluate the overall estimation accuracy of the pointwise distribution of $Y_i(t, X_i)$. Specifically, if $g(\cdot; t, x)$ denotes the true density function of $Y_i(t, x)$ for time point $t$ and covariate $x$, and $\hat{g}(\cdot; t, x)$ is the estimated density function obtained using the estimated mean, variance and shape parameter functions, the integrated KL divergence (IKL) is defined as $\text{IKL}(g) = \int_0^1 \int_0^1 \text{KL}\{g(t, x)\} dtdx$, where $\text{KL}\{g(t, x)\} = \int \hat{g}(s; t, x) \{\log \hat{g}(s; t, x) - \log g(s; t, x)\} ds$.

The prediction performance is measured by the mean prediction error (MPE) calculated on a test data set with sample size 100. For the subject $i$ in the test data, half the locations are randomly chosen to be missing, say $\{t_{i,1}, \ldots, t_{i,40}\}$, then $\text{MPE} = \sqrt{\sum_{i=1}^{100} \sum_{j=1}^{40} (\hat{Y}_i(t_{i,j}, X_i) - Y_i(t_{i,j}, X_i))^2}/4000$.

Table 1 illustrates the accuracy in estimating the model components. Firstly, it shows numerical evidence that the proposed approach for estimating the population level functions is consistent; see the decreased IMSE for the population level functions and decreased IKL as the sample size increases, as shown by the columns 3 – 5 corresponding to cSFM. The accuracy of the model dependence estimation is confirmed by the prediction error which decreases for increased sample size (see last column for $n = 100$ to 300). Secondly, it shows that cSFM compares favorably with the other alternatives. In particular, assuming the same semi-parametric model but not accounting for the time-varying skewness, affects slightly the estimation of the mean function, but has a pronounced negative effect on the estimation of the other components as well as on the prediction; compare the results for cSFM and cSFM$_0$. Joint estimation of the population level functions with regression splines (cSFM) gives more accurate estimates/prediction than using a stepwise procedure (two-step cSFM), even the penalized splines with REML-based smoothing parameters selection is applied to estimate the mean at the first step. The main competitor is mFPCA, which uses a fully non-parametric modeling framework and accounts for the covariate solely in the mean function. The prediction performance is improved by fitting the proposed semi-parametric model with a stepwise
estimation approach (two-step cSFM) and even more so by using simultaneous estimation (cSFM), which accounts for the covariate dependence and time-varying skewness.

Table 2 assesses the performance of the cSFM with other alternatives in terms of pointwise quantile function estimation. The inappropriate normality assumption made by mFPCA leads to biased quantile estimators, especially for close-to-boundaries level of quantiles, and this is noticed by the high inaccuracy of the estimates even for large sample size (see the rows when $n = 300$). In contrast, by accounting for the time-varying skewness appropriately, the estimated pointwise quantile functions are very accurate, in particular for quantile levels that are close to boundaries (e.g. 95\%th or 99\%th quantile levels); compare the performance of cSFM and two-step cSFM with the other approaches. These findings are consistent for all the sample sizes examined. As expected, since COBS makes no parametric assumption on the underlying pointwise distribution, the accuracy of the method deteriorates greatly as the level approaches the boundaries, as fewer and fewer observations are relevant for estimation. On the other hand, the quantile estimation with cSFM relies on the estimates of population-level parameter functions, which are obtained using all data, and the accuracy of the estimates is invariant to the quantile levels. The accuracy of the quantile functions estimation with cSFM depends on the parametric assumptions.

6 Analysis of the Tractography Data

Our motivating application is a brain tractography study for multiple sclerosis (MS) disease. Diffusion tensor imaging (DTI) is a magnetic resonance imaging technique that allows to visualize the white matter tracts of the brain and has been used extensively in the study of MS (Greven and others, 2010). In this paper we consider two adjacent tracts: corpus callosum (CCA), which connects the two hemispheres of the brain, and the left/right cortico-spinal tracts (lCST/rCST) which connects the left/right part of the brain with the spinal cord and focus on the DTI modality that describes the amount of water diffusivity along the direction of the tract, parallel diffusivity (LO). Our goal to describe how the parallel diffusivity varies along the CCA tract, while accounting
for the mean summary of the parallel diffusivity along the nearby tract, lCST (mean lCST-LO). Because the two tracts are in close proximity to one another it is expected that the behavior of the parallel diffusivity profile along CCA is affected by the mean lCST-LO.

The study consists of DTI measurements for 160 MS subjects and 42 healthy individuals taken at the baseline visit. For each subject, the following variables are recorded: LO measurements at 93 equidistant locations along the CCA tract, and the mean summary of the LO along the lCST. Part of the data is available in the R-package refund (Crainiceanu and others, 2012). Figure 1 displays parallel diffusivity profiles along CCA, for the corresponding covariate mean lCST-LO - plotted separately for the MS group and the healthy subjects group (controls).

We take LO along CCA as the profile of interest and the mean summary of the lCST-LO as the available continuous covariate. Separately for MS and controls, consider the proposed cSFM using Gaussian copula to model the dependence. It is assumed that the mean lCST-LO affects both the mean and the variance at each location, and that $G\{\cdot, \alpha(\cdot)\}$ is the centered and scaled skewed Normal distribution with shape parameter $\alpha(\cdot)$, that depends on the tract location solely. Tensor product of cubic B-splines with equal number of knots in each direction are used to model the population level functions; let $K_\mu$, $K_\sigma$, and $K_\alpha$ be the number of knots used in the modeling of $\mu$, $\sigma$, and $\alpha$. The optimal number of knots for $(K_\mu, K_\sigma, K_\alpha)$ as selected by AIC is (9,5,4) for the MS group, and (6,3,2) for the control group. We use bootstrap by re-sampling the pairs of the subject’s LO profile along the CCA plus its mean summary of the lCST-LO, to assess the variability of the estimators in each group. The results are presented for 1000 bootstraps samples.

Figure 2 shows the estimated mean/log-variance/skewness functions separately for the MS and control group. The first two plots reflect different range of values for the mean summary of lCST-LO for MS patients and healthy controls. Our findings confirm prior results of Staicu and others (2012) that the overall LO exhibits a surge at the beginning of the CCA tract, followed by a slight decline and then a gradual increase towards the end of the tract, and in addition it characterizes
how this behavior changes with the mean summary of the lCST-LO. The overall LO at some fixed location along the CCA tract exhibits a wavy dependence on the mean lCST-LO with multiple local moderate peaks values for both groups. For MS subjects the overall LO is decreased for values of the mean summary of the lCST-LO that are in the middle of the observed range compared to the counterparts corresponding to more extreme values of the mean summary of the lCST-LO; this is different for the control group. The rightmost plots show the estimated skewness in the two groups: using 90% pointwise bootstrap confidence intervals, we find evidence of significant pointwise skewness in the parallel diffusivity towards the end of the CCA tract, for MS patients.

These results are investigated further via bootstrap; in particular we examine whether there is significant difference between the corresponding various model components in the MS and control group. We consider the values for the mean summary of the lCST-LO that are common to both groups. Figure 3 shows the map with significant differences between MS and control groups corresponding to the pointwise quantile functions of levels 50%, 95% and 99%, using pointwise bootstrap-based confidence intervals. It appears that for higher quantile levels - 95th, 99th - the parallel diffusivity is significantly different for the MS and healthy individuals for tract locations in the second half of the CCA tract, and for mean lCST-LO levels between $1.14 \times 10^{-3}$ and $1.16 \times 10^{-3}$. By comparison, Staicu and others (2012) concluded that quantiles at 50th, 95th and 99th levels are significantly different for the two groups at almost all locations along the CCA tracts. By accounting for the additional information of the mean summary of the parallel diffusivity along the lCST tract, we gain more insight into the process that describes the variation of the parallel diffusivity along the CCA, and how it is affected by the level of the mean lCST-LO. Our results indicate that when the mean lCST-LO is about $1.17 \times 10^{-3}$ there is no significant difference in the overall parallel diffusivity profile along CCA between the two groups; reversed results hold for the case when the mean summary of the lCST-LO is about $1.17 \times 10^{-3}$. These observations call for more formal investigations of whether the covariate has a significant effect in either the mean or
the variance functions for either of the groups.

We further conduct an out-of-sample predictive study for MS subjects, which are randomly divided into a training set (95%) and a test set (5%). The training data consist of complete and noisy subject profiles and subject covariate information, while the test data consist of incomplete and noisy profiles, with the last 10% of the tract assumed unobserved and corresponding covariate information. Prediction error is determined as in Section 5 based on 200 replications; the mean prediction error is $9.37 \times 10^{-5}$. By comparison the FPCA (Yao and others, 2005) and the covariate-adjusted mFPCA (Jiang and Wang, 2010) produced higher prediction errors: $(10.25, 10.17) \times 10^{-5}$ respectively; all the three prediction errors have the maximum standard error $0.18 \times 10^{-5}$.

Supplementary Material

A supplementary appendix, available online at Biostatistics, includes 1) discussion about validation of the model assumption; 2) additional simulation results when the model is misspecified and 3) additional results for the DTI tractography application.

Acknowledgment

Staicu’s research was supported by NSF grant DMS 1007466 and NIH grant 1R01NS085211-01. Bondell’s research was supported by NSF grant DMS-1308400 and NIH grant P01-CA-142538. The authors thank Ciprian Crainiceanu, Daniel Reich, and Peter Calabresi for the DTI dataset. The content is solely the responsibility of the authors and does not necessarily represent the social views of the National Institutes of Health. The authors report no conflict of interests.

References


Table 1: Comparison among the proposed cSFM using regression splines and simultaneous fitting procedure, cSFM0 - a variant of cSFM when $\alpha(t) \equiv 0$, the two-step cSFM - a variant of cSFM using a combination of penalized splines and regression splines and stepwise fitting procedure, and mFPCA (Jiang and Wang, 2010) based on a penalized smoothing estimation procedure. Displayed are: $\sqrt{IMSE}$ for the estimators of mean/log variance/skewness functions, integrated Kullback-Leibler (IKL) divergence of the pointwise distributions, and the mean prediction errors (MPE) for various sample sizes $n$; results are multiplied by 1000. Standard errors are reported in parenthesis.

<table>
<thead>
<tr>
<th>Method</th>
<th>$n$</th>
<th>mean</th>
<th>$\sqrt{IMSE}$</th>
<th>log variance</th>
<th>skewness</th>
<th>IKL</th>
<th>MPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>cSFM</td>
<td>100</td>
<td>5.07(0.17)</td>
<td>453.82(8.30)</td>
<td>169.14(5.75)</td>
<td>71.07(4.23)</td>
<td>9.97(0.10)</td>
<td></td>
</tr>
<tr>
<td>cSFM0</td>
<td>100</td>
<td>5.00(0.17)</td>
<td>484.81(8.45)</td>
<td>826.65(0.00)</td>
<td>356.27(8.93)</td>
<td>9.97(0.06)</td>
<td></td>
</tr>
<tr>
<td>two-step cSFM</td>
<td>100</td>
<td>6.90(0.15)</td>
<td>545.37(6.88)</td>
<td>255.86(7.99)</td>
<td>224.3(13.83)</td>
<td>9.90(0.06)</td>
<td></td>
</tr>
<tr>
<td>mFPCA</td>
<td>100</td>
<td>6.90(0.15)</td>
<td>2769.03(15.38)</td>
<td>-</td>
<td>-</td>
<td>9.97(0.06)</td>
<td></td>
</tr>
<tr>
<td>cSFM</td>
<td>200</td>
<td>3.54(0.11)</td>
<td>287.81(4.96)</td>
<td>109.78(3.15)</td>
<td>30.54(0.71)</td>
<td>9.38(0.05)</td>
<td></td>
</tr>
<tr>
<td>cSFM0</td>
<td>200</td>
<td>3.73(0.11)</td>
<td>307.21(6.04)</td>
<td>826.65(0.00)</td>
<td>313.5(5.65)</td>
<td>9.80(0.06)</td>
<td></td>
</tr>
<tr>
<td>two-step cSFM</td>
<td>200</td>
<td>5.11(0.11)</td>
<td>439.94(5.94)</td>
<td>217.42(6.07)</td>
<td>120.78(4.82)</td>
<td>9.58(0.06)</td>
<td></td>
</tr>
<tr>
<td>mFPCA</td>
<td>200</td>
<td>5.11(0.11)</td>
<td>2792.02(11.87)</td>
<td>-</td>
<td>-</td>
<td>9.87(0.06)</td>
<td></td>
</tr>
<tr>
<td>cSFM</td>
<td>300</td>
<td>2.73(0.09)</td>
<td>224.64(3.95)</td>
<td>100.26(2.29)</td>
<td>20.31(0.42)</td>
<td>9.15(0.06)</td>
<td></td>
</tr>
<tr>
<td>cSFM0</td>
<td>300</td>
<td>2.87(0.09)</td>
<td>242.81(4.31)</td>
<td>826.65(0.00)</td>
<td>318.54(4.87)</td>
<td>9.57(0.06)</td>
<td></td>
</tr>
<tr>
<td>two-step cSFM</td>
<td>300</td>
<td>4.15(0.09)</td>
<td>384.92(4.97)</td>
<td>210.88(5.95)</td>
<td>98.48(5.41)</td>
<td>9.37(0.06)</td>
<td></td>
</tr>
<tr>
<td>mFPCA</td>
<td>300</td>
<td>4.15(0.09)</td>
<td>2782.63(7.97)</td>
<td>-</td>
<td>-</td>
<td>9.69(0.06)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Observed parallel diffusivity (LO) along the corpus callosum tract locations and the covariate lCST-mLO, for 160 MS and 42 healthy subjects.
Table 2: Quantile estimation performance in terms of $\sqrt{\text{IMSE}}$. Results are displayed for quantile functions estimators at levels 50%, 80%, 90%, 95% and 99% obtained with cSFM, cSFM$_0$, two-step cSFM, mFPCA, and COBS for various sample sizes $n$; results are multiplied by 1000. Standard errors are reported in parenthesis.

<table>
<thead>
<tr>
<th>Method</th>
<th>$n$</th>
<th>50%</th>
<th>80%</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>cSFM</td>
<td>100</td>
<td>4.81(0.17)</td>
<td>5.81(0.18)</td>
<td>6.98(0.22)</td>
<td>8.16(0.26)</td>
<td>10.83(0.34)</td>
</tr>
<tr>
<td>cSFM$_0$</td>
<td>100</td>
<td>6.37(0.16)</td>
<td>5.95(0.20)</td>
<td>7.54(0.23)</td>
<td>10.51(0.26)</td>
<td>18.27(0.33)</td>
</tr>
<tr>
<td>two-step cSFM</td>
<td>100</td>
<td>6.71(0.15)</td>
<td>7.60(0.15)</td>
<td>8.75(0.19)</td>
<td>9.98(0.23)</td>
<td>12.78(0.31)</td>
</tr>
<tr>
<td>mFPCA</td>
<td>100</td>
<td>7.98(0.15)</td>
<td>14.65(0.15)</td>
<td>21.06(0.18)</td>
<td>27.07(0.20)</td>
<td>39.45(0.25)</td>
</tr>
<tr>
<td>COBS</td>
<td>100</td>
<td>7.28(0.13)</td>
<td>9.19(0.17)</td>
<td>11.32(0.25)</td>
<td>13.44(0.27)</td>
<td>20.89(0.34)</td>
</tr>
<tr>
<td>cSFM</td>
<td>200</td>
<td>3.43(0.1)</td>
<td>4.10(0.14)</td>
<td>4.92(0.17)</td>
<td>5.77(0.19)</td>
<td>7.66(0.25)</td>
</tr>
<tr>
<td>cSFM$_0$</td>
<td>200</td>
<td>5.5(0.14)</td>
<td>4.42(0.14)</td>
<td>5.91(0.14)</td>
<td>9.04(0.17)</td>
<td>17.03(0.23)</td>
</tr>
<tr>
<td>two-step cSFM</td>
<td>200</td>
<td>4.98(0.11)</td>
<td>5.61(0.12)</td>
<td>6.44(0.14)</td>
<td>7.33(0.16)</td>
<td>9.37(0.21)</td>
</tr>
<tr>
<td>mFPCA</td>
<td>200</td>
<td>6.52(0.13)</td>
<td>13.98(0.10)</td>
<td>20.69(0.11)</td>
<td>26.87(0.13)</td>
<td>39.49(0.17)</td>
</tr>
<tr>
<td>COBS</td>
<td>200</td>
<td>4.76(0.08)</td>
<td>5.61(0.1)</td>
<td>7.12(0.12)</td>
<td>9.37(0.15)</td>
<td>17.59(0.28)</td>
</tr>
<tr>
<td>cSFM</td>
<td>300</td>
<td>2.67(0.08)</td>
<td>3.16(0.10)</td>
<td>3.80(0.12)</td>
<td>4.51(0.14)</td>
<td>6.18(0.18)</td>
</tr>
<tr>
<td>cSFM$_0$</td>
<td>300</td>
<td>4.37(0.11)</td>
<td>3.49(0.11)</td>
<td>4.93(0.11)</td>
<td>8.20(0.13)</td>
<td>16.26(0.18)</td>
</tr>
<tr>
<td>two-step cSFM</td>
<td>300</td>
<td>4.09(0.09)</td>
<td>4.53(0.10)</td>
<td>5.14(0.12)</td>
<td>5.84(0.13)</td>
<td>7.55(0.17)</td>
</tr>
<tr>
<td>mFPCA</td>
<td>300</td>
<td>5.80(0.11)</td>
<td>13.78(0.09)</td>
<td>20.54(0.10)</td>
<td>26.73(0.11)</td>
<td>39.31(0.14)</td>
</tr>
<tr>
<td>COBS</td>
<td>300</td>
<td>4.04(0.07)</td>
<td>4.82(0.08)</td>
<td>6.35(0.11)</td>
<td>8.89(0.15)</td>
<td>17.13(0.22)</td>
</tr>
</tbody>
</table>
Figure 2: Estimated mean/log-variance surfaces and skewness functions for MS (top panels) and Control group (bottom). Pointwise 90% confidence intervals (dotted line) from 1000 bootstrap replicates are constructed for the estimated skewness function in the two groups.
Figure 3: Significance maps for the difference between the MS and control based on 1000 bootstraps. Results are shown for the mean/log variance/skewness (top panels) and quantile functions at various quantile levels (bottom). The color coding is “light grey” for not significant differences, “red” significant negative differences, and “blue” significant positive differences. Pointwise 90% confidence intervals from 1000 bootstrap replicates are used to measure the significance.