ORIGINAL ARTICLE

Conditional distribution regression for functional responses

Jianing Fan🟢 | Hans-Georg Müller

Department of Statistics, University of California, Davis, California

Correspondence
Jianing Fan, Department of Statistics, University of California, Davis, CA, USA.
Email: jngfan@ucdavis.edu

Abstract

Modeling conditional distributions for functional data extends the concept of a mean response in functional regression settings, where vector predictors are paired with functional responses. This extension is challenging because of the nonexistence of well-defined densities, cumulative distributions, or quantile functions in the Hilbert space where the response functions are located. To address this challenge, we simplify the problem by assuming that the response functions are Gaussian processes, which means that the conditional distribution of the responses is determined by conditional mean and conditional covariance. We demonstrate that these quantities can be obtained by applying global and local Fréchet regression, where the local version is more flexible and applicable when the covariate dimension is low and covariates are continuous, while the global version is not subject to these restrictions but is based on the assumption of a more restrictive regression relation. Convergence rates for the proposed estimates are obtained under the framework of M-estimation. The corresponding estimation of conditional distributions is illustrated with simulations and an application to bike-sharing data, where predictors include weather characteristics and responses are bike rental profiles. We also show that our methods are applicable to the challenging problem to study functional fragments. Such

data are observed in accelerated longitudinal studies and correspond to functional data observed over short domain segments. We demonstrate the utility of conditional distributions in this context by using the time (age) at which a subject enters the domain of a fragment in addition to other covariates as predictor and the function observed over the domain of the fragment as response.

1 | INTRODUCTION

In functional data analysis, regression models for the case where predictors are vectors and responses are functions have been well developed and there are multiple approaches available for this function-on-vector regression setting. A widely used model is functional linear regression (Chiou et al., 2003, 2004; Faraway, 1997; Ramsay & Dalzell, 1991). Research to date has mainly focused on the mean regression $E(Y|X=x)$. For the case of scalar responses Y it is well known that often relevant information can be obtained by studying and modeling conditional distributions $P(Y \leq y|X=x)$ or the equivalent quantile regression (Koenker & Bassett Jr, 1978), rather than just considering the first conditional moment $E(Y|X=x)$. It is a challenge to extend these notions to the case where responses are functions, as quantiles cannot be naturally defined in the infinite-dimensional Hilbert space L^2, where functional responses are typically assumed to live. We demonstrate that one can gain more information in functional regression settings by moving beyond mean regression to conditional distribution modeling when responses are random functions, and develop an approach that is supported by theory and can be easily implemented.

Specifically, we consider pairs of random variables (X, Y) of functional responses $Y \in L^2(\mathcal{T})$, where the domain $\mathcal{T}$ is an interval, and vector predictors $X \in \mathcal{R}^p$ and assume that data samples are available from the joint distribution $(X, Y) \sim F$. The object of interest is the conditional distribution $\mathcal{L}(Y|X)$. For scalar responses, when $Y \in \mathcal{R}$, $\mathcal{L}(Y|X)$ can be quantified as a conditional quantile function or a conditional cumulative distribution function (CDF). This problem has been studied widely, even for the case of functional or time-varying predictors (Chen & Müller, 2012; Ding et al., 2018; Kato, 2012; Wang et al., 2009; Yao et al., 2017). However, the case of functional responses is much more challenging, as it is more complex than the related problem of finding conditional distributions for multivariate responses, due to the fact that the dimension of the function space is infinity. It is therefore not surprising that for functional Y the modeling and study of conditional distributions $\mathcal{L}(Y|X)$ has remained largely unexplored.

When Y is functional, the modeling of conditional spatial depth, which is related to conditional spatial distributions, has been investigated by Chowdhury and Chaudhuri (2016), but

the only previous work that directly addresses the conditional distribution problem that we are aware of is Chen and Müller (2014), where an estimator was introduced under the assumption that $\mathcal{L}(Y|X)$ is Gaussian, the functional principal components of Y are assumed to be conditionally uncorrelated, that is, $\text{Cov}(\xi_k, \xi_l|X) = 0$ and the eigenfunctions ψ_k of the conditional covariance operator do not depend on the predictor X. While this estimator was heuristically motivated and worked well in an application to traffic data, no supporting theory was provided.

Assuming that the conditional distribution of interest $\mathcal{L}(Y|X)$ is Gaussian, it is determined by the conditional mean function and the conditional covariance surface. While conditional mean modeling for functional responses has been well studied (Chiou et al., 2004; Faraway, 1997; Morris, 2015; Ramsay, 2006; Wang et al., 2016), considering nonnegative definite covariance surfaces as responses is a novel challenge that has not been well studied before, in contrast to the case of covariance matrices as responses (Petersen et al., 2019). For example, when considering linear models, the nonlinear constraints inherent in covariance surfaces are generally not satisfied for the fitted responses. In previous work, kernel-type methods have been considered for estimating conditional covariance surfaces when predictors are scalar or vectors (Cardot, 2007; Chiou & Müller, 2009; Jiang & Wang, 2010). Such methods can realistically only be applied when the dimension of the predictors is small due to the curse of dimensionality and do not ensure nonnegative definiteness of the covariance surface estimates.

We avoid these problems by applying local and global versions of Fréchet regression (Petersen & Müller, 2019). When choosing the Frobenius metric for the case of covariance surfaces, the target coincides with the usual conditional covariance surface $\text{Cov}(Y(s), Y(t)|X)$. Local Fréchet regression is a flexible approach when the predictors X are low dimensional and the regression relation is smooth, while global Fréchet regression does not require a tuning parameter and is applicable for multivariate predictors.

In addition to conditional distributions of bike rental daily profiles in response to weather and other conditions, which we report in Section 5, an intriguing application of our approach is to the functional fragment problem. This challenging problem arises in accelerated longitudinal studies, where only segments or snippets of functional data are observed over time domains that are short relative to the entire time domain, which is a consequence of the short time span over which such studies are conducted. At issue are the conclusions that can be reasonably drawn from such sparse data, where covariance operators cannot be reliably estimated and thus eigenanalysis of the functional data is not feasible. For instance, in an accelerated study of growth one might observe body length for each subject only during a randomly located fixed length time period for each child, such as a period of a few years, instead of observing the child over the entire time domain of 20 years of growth. We propose a new approach to deal with functional fragments by using the time (age) at entry into the domain of a fragment and possibly other covariates as predictors and the function observed over the time domain of the fragment as response. We can then determine the time-varying conditional distribution of the functional fragments to describe the time-evolution of the functional fragments.

The article is organized as follows. In Section 2 global and local versions of Fréchet regression are reviewed with a view toward conditional mean and covariance estimation. We then demonstrate the modeling of conditional functional distributions and derive simultaneous prediction bands. In Section 3 we study the asymptotic properties of the estimates. Details of implementation and simulation results are in Section 4, while data analysis illustrations can be found in Section 5. Proofs and auxiliary results are collected in the Appendix (Online Supplement).

2 | CONDITIONAL MEAN AND COVARIANCE ESTIMATION

2.1 | Background on Fréchet regression

Since the space of nonnegative definite covariance surfaces is not a linear space, traditional regression models are not directly extendable to this case. Fréchet regression is a generalization of Fréchet means that were introduced by Fréchet (1948) as an extension of the notion of a mean to a general metric space. The global and local models in Fréchet regression are extensions of linear and local linear regression. In fact, when the metric space is a Hilbert space, these models are equivalent to linear and local linear regression, as demonstrated in the conditional mean estimation section below. For more details about some of the developments that are briefly reviewed below we refer to Petersen and Müller (2019). For random objects Z taking values in a bounded metric space Ω with metric d, Fréchet means are defined as

$$\omega_\oplus = \arg\min_{\omega \in \Omega} E(d^2(Z, \omega)),$$

where their existence and uniqueness depends on structural properties of the underlying metric space.

Considering $(X, Z) \sim F$, where X and Z take values in R^p and Ω, respectively, assume that the mean $\mu = E(X)$, the covariance matrix $\Sigma = \text{Var}(X)$ of X and the conditional distributions $\mathcal{L}(X|Z)$ and $\mathcal{L}(Z|X)$ are well defined. The Fréchet regression of Z, given $X = x$, targets

$$m_\oplus(x) = \arg\min_{\omega \in \Omega} M_\oplus(\omega, x), \quad M_\oplus(\cdot, x) = E(d^2(Z, \cdot)|X = x),$$

where we assume existence and uniqueness of these quantities.

The global Fréchet regression model is defined as

$$\tilde{m}_\oplus(x) = \arg\min_{\omega \in \Omega} \tilde{M}_\oplus(\omega, x), \quad \tilde{M}_\oplus(\cdot, x) = E[w_G(x; X)d^2(Z, \cdot)],$$

where the weight function w_G is given by $w_G(x; X) = 1 + [X - \mu]^T \Sigma^{-1}(x - \mu)$, $\mu = E(X)$, $\Sigma = \text{Var}(X)$, and corresponding M-estimators are given by

$$\hat{m}_\oplus(x) = \arg\min_{\omega \in \Omega} \tilde{M}_n(\omega, x), \quad \tilde{M}_n(\cdot, x) = n^{-1} \sum_{i=1}^n w_{iG}(x)d^2(Z_i, \cdot). \tag{1}$$

with weights $w_{iG}(x) = 1 + (X_i - \overline{X})^T \hat{\Sigma}^{-1}(x - \overline{X})$.

For local Fréchet regression we present results for the case of scalar predictors, where the extension to multivariate predictors is straightforward. With a nonnegative kernel function K that integrates to 1 and using the notation $K_h(\cdot) = \frac{1}{h}K(\frac{\cdot}{h})$ for a kernel scaled with bandwidth h, the local Fréchet regression model can be written as

$$\tilde{l}_\oplus(x) = \arg\min_{\omega \in \Omega} L(\omega, x), \quad L(\cdot, x) = E(w_L(x, h; X)d^2(Z, \cdot))$$

with $w_L(x, h; X) = \frac{1}{\sigma_0^2} K_h(X - x)[\mu_2 - \mu_1(X_i - x)]$, where $\sigma_0^2 = \mu_0\mu_2 - \mu_1^2$ and $\mu_j = E(K_h(X - x)(X - x)^j)$.

The corresponding local Fréchet estimator is

$$\hat{l}_\oplus(x) = \arg\min_{\omega\in\Omega} L_n(\omega, x), \quad L_n(\cdot, x) = n^{-1}\sum_{i=1}^n w_{iL}(x, h)d^2(Z_i, \cdot), \tag{2}$$

employing weights $w_{iL}(x, h) = \frac{1}{\hat\sigma_0^2}K_h(X_i - x)[\hat\mu_2 - \hat\mu_1(X_i - x)]$, where $\hat\sigma_0^2 = \hat\mu_0\hat\mu_2 - \hat\mu_1^2$ and $\hat\mu_j = n^{-1}\sum_{i=1}^n K_h(X_i - x)(X_i - x)^j$.

2.2 | Conditional mean estimation

Consider responses that take values in the Hilbert space $L^2(\mathcal{T})$. When choosing the L^2 metric in the bounded target space of mean functions $\Omega_M \subset L^2(\mathcal{T})$, it is easy to see that $E(Y(t)|X) = E_\oplus(Y|X)(t)$, where $E_\oplus$ is the Fréchet mean. Then estimation is straightforward within the Fréchet regression framework by taking $Y_i(\cdot)$, $i = 1, 2, \ldots, n$, as responses. The following result provides a novel characterization of Fréchet regression and shows that, interestingly, global and local Fréchet regression in this setting are equivalent to fitting time-varying linear or local linear regression models. We use here the notations

$$(\hat\beta_0(t), \hat\beta_1(t)) = \arg\min_{\beta_0(t),\beta(t)} \sum_{i=1}^n \left(Y_i(t) - \beta_0(t) - \beta_1^T(t)X_i\right)^2, \tag{3}$$

$$\left(\hat\beta_0^*(x, h_1, t), \hat\beta_1^*(x, h_1, t)\right) = \arg\min_{\beta_0(t),\beta_1(t)} \sum_{i=1}^n K_{h_1}(X_i - x)(Y_i(t) - \beta_0(t) - \beta_1(t)(X_i - x))^2. \tag{4}$$

Proposition 1. *Consider global and local Fréchet regression estimates with weights w_{iG} and w_{iL} defined in (1), (2), given by*

$$\hat\mu_G(x, \cdot) = \arg\min_{\beta_0} \sum_{i=1}^n w_{iG}(x)d^2(Y_i, y), \tag{5}$$

$$\hat\mu_L(x, h_1, \cdot) = \arg\min_{\beta_0} \sum_{i=1}^n w_{iL}(x, h_1)d^2(Y_i, y). \tag{6}$$

For the L^2 metric $d_2(Y_1, Y_2) = \left[\int_{\mathcal{T}}(Y_1(t) - Y_2(t))^2 dt\right]^{1/2}$ for $Y_1, Y_2 \in L^2(\mathcal{T})$, it holds that

$$\hat\mu_G(x, t) = \hat\beta_0(t) + \hat\beta_1(t)x,$$
$$\hat\mu_L(x, h_1, t) = \hat\beta_0^*(x, h_1, t).$$

This result demonstrates that with Hilbert space valued responses, global Fréchet regression fits can be interpreted as fitting a linear varying coefficient model, while local Fréchet regression can be interpreted as fitting a nonparametric varying coefficient model, providing new insights into the workings of the general concept of Fréchet regression for this important special case.

2.3 | Conditional covariance estimation

To obtain conditional covariance surfaces with the Fréchet regression framework, the first step is to represent the conditional covariance as the conditional Fréchet mean of random objects

that are constructed from the initial observations $\{Y_i(\cdot), X_i\}$, where the random objects must take values in the target space $\Omega_C \subset \{\rho(s, t) : \mathcal{T} \times \mathcal{T} \to \mathcal{R},\ \rho$ is symmetric and nonnegative definite$\}$, equipped with a suitable metric. For the random objects, we take the raw covariance surfaces $Z = (Y(t) - \mu(X, t)) \otimes (Y(s) - \mu(X, s))$ and choose the Frobenius metric $d_F^2(Z_1, Z_2) = \int_{\mathcal{T}} \int_{\mathcal{T}} (Z_1(s, t) - Z_2(s, t))^2\ ds dt$, defining the object space (Ω_C, d_F).

We find that $m_\oplus(x) = \arg\min_{\omega \in \Omega} E(d^2(Z, \omega) | X = x) = \text{Cov}(Y(s), Y(t) | X = x)$, enabling us to apply Fréchet regression to the sample $Z_i(t, s) = (Y_i(t) - \mu(X_i, t)) \otimes (Y_i(s) - \mu(X_i, s)),\ i = 1, 2, \dots, n$. However, since $\mu(x, t)$ is unknown, analogous to the situation considered in Petersen et al. (2019) for the case of covariance matrices, the sample elements Z_i are not directly available and proxies need to be obtained by estimating the conditional mean by $\hat{\mu}_{iG}(x, t)$ or $\hat{\mu}_{iL}(x, t)$, substituting these estimates for the unknown true means. Then the available sample consists of the tensor products

$$Z_{iG}(t, s) = (Y_i(t) - \hat{\mu}_G(X_i, t)) \otimes (Y_i(s) - \hat{\mu}_G(X_i, s)), \tag{7}$$

$$Z_{iL, h_1}(t, s) = (Y_i(t) - \hat{\mu}_L(X_i, h_1, t)) \otimes (Y_i(s) - \hat{\mu}_L(X_i, h_1, s)), \tag{8}$$

for global, respectively, local Fréchet regression. We show in Section 3 that there is no loss when substituting Z_{iL}, Z_{iG} for the unknown true Z_i, under mild regularity assumptions.

The global and local versions of the Fréchet regression for covariance surfaces, that is, the conditional covariance estimates, are then given by

$$\hat{C}_G(x, \cdot, \cdot) = \arg\min_{\omega \in \Omega_C} n^{-1} \sum_{i=1}^n w_{iG}(x) d_F^2(Z_{iG}, \omega), \tag{9}$$

$$\hat{C}_{L, h_1, h_2}(x, \cdot, \cdot) = \arg\min_{\omega \in \Omega_C} n^{-1} \sum_{i=1}^n w_{iL}(x, h_2) d_F^2(Z_{iL, h_1}, \omega). \tag{10}$$

Let

$$\tilde{C}_G(x, t, s) = n^{-1} \sum_{i=1}^n w_{iG}(x) Z_{iG}, \quad \tilde{C}_{L, h_1, h_2}(x, t, s) = n^{-1} \sum_{i=1}^n w_{iL}(x, h_2) Z_{iL, h_1}$$

with eigenvalue decomposition

$$\tilde{C}_G(x, t, s) = \sum_{k=1}^\infty \lambda_{kG}(x) \psi_{kG}(x, s) \psi_{kG}(x, t), \tag{11}$$

$$\tilde{C}_{L, h_1, h_2}(x, t, s) = \sum_{k=1}^\infty \lambda_{kL}(x) \psi_{kL}(x, s) \psi_{kL}(x, t). \tag{12}$$

Then the explicit solutions can be expressed as

$$\hat{C}_G(x, t, s) = \sum_{\lambda_{kG} > 0} \lambda_{kG}(x) \psi_{kG}(x, s) \psi_{kG}(x, t),$$

$$\hat{C}_{L, h_1, h_2}(x, t, s) = \sum_{\lambda_{kL} > 0} \lambda_{kL}(x) \psi_{kL}(x, s) \psi_{kL}(x, t),$$

where further details can be found in the proofs of Propositions 3 and 5.

2.4 | Prediction bands

Under Gaussian assumptions, estimates for the conditional mean and covariance functions yield the conditional distribution $\mathcal{L}(Y|X)$, as it is uniquely determined by the conditional mean function and the conditional covariance surface. This approach can also be utilized to construct an approximate simultaneous prediction band of $Y|X = x$. Decomposing the conditional covariance surface $C(x, s, t)$ into the orthonormal eigenfunctions of the conditional covariance operator, by Mercer's theorem,

$$C(x, s, t) = \sum_{k=1}^{\infty} \lambda_k(x)\psi_k(x, s)\psi_k(x, t).$$

Assuming that the conditional eigenfunctions $\psi_k(x, \cdot)$ form a basis of $L^2(\mathcal{T})$ for each x, we may represent $Y|X = x$ in this basis,

$$Y(t) = \mu(x, t) + \sum_{k=1}^{\infty} \xi_k(x)\psi_k(x, t). \tag{13}$$

The variance of the conditional functional principal components is $\text{Var}(\xi_k|X = x) = \lambda_k(x)$, where $\lambda_k(x) > 0$ is the kth eigenvalue of the conditional covariance operator C_x. Instead of using infinitely many components, for practical purposes we need to truncate the sum in (13) at J_n included summands, which we choose as the smallest integer satisfying

$$\inf_x \frac{\sum_{k=1}^{J_n} \lambda_k(x)}{\sum_{k=1}^{\infty} \lambda_k(x)} \geq r_n, \tag{14}$$

where r_n is the fraction of variance explained. By letting $r_n \to 1$ as $n \to \infty$, one can eventually recover the entire conditional distribution.

Assuming conditional Gaussianity, that is, that $\{\xi_1(x), \ldots, \xi_{J_n}(x)\}|X = x$ is also Gaussian distributed for all x, one can construct $(1 - \alpha)$ prediction regions via a simple ellipsoid construction based on

$$P\left\{\sum_{k=1}^{J_n} \frac{\xi_k(x)^2}{\lambda_k(x)} > \chi^2_{J_n, 1-\alpha}\right\} \leq \alpha.$$

From this one obtains the $(1 - \alpha)$ level simultaneous (in $t \in \mathcal{T}$) prediction band

$$\mathcal{B}(1 - \alpha, x, r_n) = \left\{\mu(x, t) + \sum_{k=1}^{J_n} a_k \lambda_k^{1/2}(x)\psi_k(x, t) : \sum_{k=1}^{J_n} a_k^2 \leq \chi^2_{J_n, 1-\alpha}, \quad t \in \mathcal{T}\right\},$$

denoting by $\chi^2_{J_n, 1-\alpha}$ the $(1 - \alpha)$ quantile of the χ^2 distribution with J_n degrees of freedom. This construction involves an approximation, due to the truncation at J_n included terms. We will show in Section 3 that these prediction bands achieve correct coverage in the limit when $n \to \infty$ and $J_n \to \infty$.

In applications the conditional mean and covariance need to be estimated, for which we use the methods of Section 2.2. Specifically, we use global estimates $(\hat{\mu}_G(x, \cdot), \hat{C}_G(x, \cdot, \cdot))$ (5), (6) or

local estimates $(\hat{\mu}_{L,h_1}(x, \cdot), \hat{C}_{L,h_1,h_2}(x, \cdot, \cdot))$ (9), (10) to replace $C(x, s, t)$ and then obtain the estimated simultaneous prediction bands

$$\mathcal{B}_G(1 - \alpha, x, J_n) = \left\{ \hat{\mu}_G(x, t) + \sum_{k=1}^{J_n} a_k \lambda_{kG}^{1/2}(x) \psi_{kG}(x, t) : \sum_{k=1}^{J_n} a_k^2 \leq \chi_{J_n, 1-\alpha}^2, t \in \mathcal{T} \right\}, \qquad (15)$$

$$\mathcal{B}_{L,h_1,h_2}(1 - \alpha, x, J_n) = \left\{ \hat{\mu}_{L,h_1}(x, t) + \sum_{k=1}^{J_n} a_k \lambda_{kL}^{1/2}(x) \psi_{kL}(x, t) : \sum_{k=1}^{J_n} a_k^2 \leq \chi_{J_n, 1-\alpha}^2, t \in \mathcal{T} \right\}, \quad (16)$$

where $\lambda_{kG}(x), \psi_{kG}(x, s), \lambda_{kL}(x), \psi_{kL}(x, s)$ are as in (11).

These prediction bands will be illustrated in simulations in Section 4 and for applications in Section 5.

3 | ASYMPTOTIC PROPERTIES

We establish consistency of the proposed estimates under the framework of M-estimation. A function $f : \mathcal{X} \rightarrow R$ is called α differentiable if it has uniformly bounded partial derivatives of order $\lfloor \alpha \rfloor$ (the greatest integer smaller than α) and its highest partial derivatives are Lipschitz of order $\alpha - \lfloor \alpha \rfloor$.

For $k = (k_1, k_2, \ldots, k_d)$, $k_. = \sum_{i=1}^d k_i$ and the corresponding partial derivative $D^k f$, define

$$\|f\|_\alpha = \max_{k_. \leq \lfloor \alpha \rfloor} \sup_{x \in A} |D^k f(x)| + \max_{k_. = \lfloor \alpha \rfloor} \sup_{x,y \in A} \frac{|D^k f(x) - D^k f(y)|}{\|x - y\|^{\alpha - \lfloor \alpha \rfloor}}.$$

Then $C_D^\alpha(\mathcal{X})$ denotes the set of functions that are almost everywhere α differentiable with $\|f\|_\alpha \leq D$ (see section 7, van der Vaart & Wellner, 1996). We require the following assumptions:

(A0) T is compact and $\mathcal{L}(Y|X = x)$ is Gaussian for all x.

(A1) $\Omega_M \subset C_D^\alpha(\mathcal{T})$, $\Omega_C \subset C_D^\alpha(\mathcal{T} \times \mathcal{T})$ and $Y(\cdot)$ is in $C_D^\alpha(\mathcal{T})$ with probability 1, for some $\alpha > 1$ and $D > 0$, where Ω_M is the target space of conditional mean functions and Ω_C is the target space of conditional covariance functions, as defined in Section 2.3.

(K0) The kernel K used in local Fréchet regression is a symmetric probability density function, such that with $K_{jm} = \int_R K^j(u) u^m du$, $|K_{14}|$ and $|K_{26}|$ are both finite.

The Gaussianity assumption is needed to make the problem of assessing conditional distributions in the infinite-dimensional setting tractable, as explained above. The assumption on the differentiability of the random trajectories and conditional mean function is in line with typical smoothness assumptions in functional regression.

To state our main results, we need to identify the targeted conditional mean and covariance functions. Extending least squares to metric data, global Fréchet regression aims at the best linear model, replacing least squares with more general distances, while local Fréchet regression provides an analogous local version, generalizing local least squares smoothing, as discussed in more detail in Section 2. The global Fréchet regression targets for mean and covariance are

$$\tilde{\mu}(x, \cdot) = \underset{\omega \in \Omega_M}{\arg \min} \, E[w_G(x; X) d^2(Y(\cdot), \omega)], \qquad (17)$$

$$\tilde{C}(x, \cdot, \cdot) = \underset{\omega \in \Omega_C}{\arg \min} \, E[w_G(x; X) d_F^2(C^*(\cdot, \cdot), \omega)] \qquad (18)$$

where $C^*(s, t) = (Y(t) - \mu(X, t)) \otimes (Y(s) - \mu(X, s))$, for the population targets $\mu(x, t) = E(Y(t)|x)$ and $C(x, s, t) = \mathrm{Var}(Y(t), Y(s)|x)$, where from now on we write $C(x)$ to denote $C(x, \cdot, \cdot)$.

Proposition 2. *For $\hat{\mu}_G(x, t)$ as defined in (5), under (A0) and (A1),*

$$\int_{\mathcal{T}} (\hat{\mu}_G(x, t) - \tilde{\mu}(x, t))^2 dt = O_p(n^{-1}),$$

where $\tilde{\mu}(x, \cdot)$ is the global Fréchet regression of the mean function (17).

An analogous consistency result holds for global Fréchet regression for covariance surfaces, where d_F denotes the Frobenius metric, as before.

Proposition 3. *For $\hat{C}_G(x, \cdot, \cdot)$ as defined in (9), under (A0) and (A1),*

$$d_F(\hat{C}_G(x), \tilde{C}(x)) = O_p(n^{-1/2}),$$

where $\tilde{C}(x)$ is the global Fréchet regression of covariance (18).

We utilize these consistency results to obtain the rate of convergence of the 2-Wasserstein distance between estimated and true conditional distributions. Since it has become one of the most popular metrics in the space of distributions, due to its role in optimal transport and its superior performance in applications (Petersen & Müller, 2016; Villani, 2008), we quantify here the distance of distributions with the Wasserstein metric, which is also motivated by the connection of this metric with the L^2 distance of quantile functions for the case of one-dimensional distributions that we consider here. The 2-Wasserstein distance between two distributions $\mathcal{L}_1, \mathcal{L}_2$ on a measurable metric space $\mathcal{X}$ with metric $d_{\mathcal{X}}$ is defined as

$$d_w^2(\mathcal{L}_1, \mathcal{L}_2) = \inf_{\tilde{X}_1 \sim \mathcal{L}_1, \tilde{X}_2 \sim \mathcal{L}_2} E[d_{\mathcal{X}}(\tilde{X}_1, \tilde{X}_2)^2]. \tag{19}$$

In our case the metric space $\mathcal{X}$ is the Hilbert space of square integrable functions with the L^2 metric. When analyzing the minimization problem (19), we will make heavy use of the fact that only Gaussian distributions $\mathcal{L}_1, \mathcal{L}_2$ need to be considered. The following main result provides the rate of convergence in terms of the Wasserstein distance of the estimated conditional distribution $\mathcal{G}(\hat{\mu}_G(x, t), \hat{C}_G(x, s, t))$ to the true conditional distribution $\mathcal{L}(Y|X)$, where $\mathcal{G}(\mu(x, t), C(x, s, t))$ denotes the Gaussian measure on $L^2(\mathcal{T})$ with conditional mean function $\mu(x, \cdot)$ at $X = x$ and conditional covariance function $C(x, \cdot, \cdot)$ at $X = x$, so that $\mathcal{L}(Y|X = x) \equiv \mathcal{G}(\mu(x, t), C(x, s, t))$.

Theorem 1. *For $\hat{\mu}_G(x, t)$ defined in (5) and $\hat{C}_G(x, \cdot, \cdot)$ defined in (9), if (A0) holds and (A1) is satisfied with $\alpha > 2$, and if $\mu(x, t) = \tilde{\mu}(x, t)$ and $C(x) = \tilde{C}(x)$, that is, the true conditional mean and covariance satisfy the linear regression models implied by global Fréchet regression, then*

$$d_w(\mathcal{L}(Y|X = x), \mathcal{G}(\hat{\mu}_G(x, t), \hat{C}_G(x, s, t))) = O_p(n^{-(\alpha-2)/8\alpha}).$$

We note that the rate of convergence in this distributional consistency result improves as the degree of smoothness α in assumption (A1) increases and approaches $n^{-1/8}$ as $\alpha \to \infty$.

A second option is to employ local Fréchet regression estimators for the conditional mean functions $\hat{\mu}_L(x)$ (6) and the conditional covariance functions $\hat{C}_L(x)$ (10). This approach is more flexible as one does not need to require that the true conditional mean and covariance functions

follow the linear model that is implied by global Fréchet regression. However, as in the Euclidean case, in exchange for fewer assumptions, in the nonparametric approach one needs to consider bias, and the rates of convergence for the conditional distributions are slower. This is seen in Theorem 2, the proof of which makes use of the following two results.

Proposition 4. *Under (A0), (A1), and (K0), if the bandwidth sequence satisfies $h_1 \sim n^{-1/5}$, it holds that*

$$d_2(\hat{\mu}_L(x, h_1, t), \mu(x, t)) = O_p(n^{-2/5}).$$

Proposition 5. *Under (A0), (A1), (K0) and $h_1 \sim h_2$*

$$d_F(\hat{C}_{L,h_1,h_2}(x), C_{\oplus}(x)) = O_p((nh_1)^{-1/2}), \quad d_F(C_{\oplus}(x), C(x)) = O_p(h_1^2),$$

where $C_{\oplus}(x, \cdot, \cdot) = argmin_{\omega \in \Omega_C} E[w_L(x, h_2; X) d_2^2(Z, \omega)]$ with $Z = (Y(t) - \mu(X, t)) \otimes (Y(s) - \mu(X, s))$ is the smoothed target and $C(x, \cdot, \cdot) = \text{Cov}(Y(\cdot), Y(\cdot) | X = x)$ is the true conditional covariance function. Then choosing $h_1, h_2 \sim n^{-1/5}$, it holds that

$$d_F(\hat{C}_{L,h_1,h_2}(x), C(x)) = O_p(n^{-2/5}).$$

The preceding results lead to the rate of convergence for the local estimation approach when targeting the true conditional distribution estimation.

Theorem 2. *Under (A0) and (K0), if (A1) is satisfied with $\alpha > 2$ and if $h_1 \sim h_2 \sim n^{-1/5}$,*

$$d_w(\mathcal{L}(Y|X = x), \mathcal{G}(\hat{\mu}_L(x, h_1, t), \hat{C}_{L,h_1,h_2}(x, s, t))) = O_p(n^{-(\alpha-2)/10\alpha}).$$

This implies that for the local approach, the convergence of the Wasserstein distance between estimated and true conditional distributions approaches the rate $n^{-1/10}$ as $\alpha \to \infty$. The above convergence of global and local Fréchet estimates of mean and covariance functions also leads to the consistency of the $1 - \alpha$ prediction bands defined in (15) and (16). The following result shows that for large sample sizes, the coverage of these bands is at least at the nominal level $1 - \alpha$.

Theorem 3. *If (A0), (A1), and (K0) holds. For the global (15) and local (16) prediction bands, as the fraction of variance explained r_n in (14) satisfies $r_n \to 1$ as $n \to \infty$,*

$$\lim_{n \to \infty} P(Y(t) \in \mathcal{B}_L(1 - \alpha, x, r_n), \quad \text{for all } t \in \mathcal{T} | X = x) \geq 1 - \alpha.$$

Furthermore, if $\tilde{\mu}(x, \cdot) = \mu(x, \cdot)$ and $\tilde{C}(x) = C(x)$,

$$\lim_{n \to \infty} P(Y(t) \in \mathcal{B}_G(1 - \alpha, x, r_n), \quad \text{for all } t \in \mathcal{T} | X = x) \geq 1 - \alpha.$$

4 | IMPLEMENTATION AND SIMULATIONS

In practical applications, one often has a dense and regular grid t_j on which the functional data $\{X_i, Y_i(t_j)\}, i = 1, 2, \ldots, n, j = 1, 2, \ldots, m$, are observed, where $\{t_1, t_2, \ldots, t_m\}$ is a dense equidistant

grid that covers $\mathcal{T}$. We then approximate the L^2 distance in function space by $\tilde{d}_2^2(Y_1, Y_2) = \frac{1}{m^2} \sum_{j=1}^m (Y_1(t_j) - Y_2(t_j))^2$ and use discretized mean function estimates for global and local versions, respectively, given by

$$\{\hat{\mu}_G(x, t_1), \hat{\mu}_G(x, t_2), \dots, \hat{\mu}_G(x, t_m)\} = \underset{y(t_j), j=1, \dots, m}{\arg\min} \sum_{i=1}^n [w_{iG}(x)\tilde{d}_1^2(Y_i, y)],$$

$$\{\hat{\mu}_L(x, h_1, t_1), \hat{\mu}_L(x, h_1, t_2), \dots, \hat{\mu}_L(x, h_1, t_m)\} = \underset{y(t_j), j=1, \dots, m}{\arg\min} \sum_{i=1}^n [w_{iL}(x, h_1)\tilde{d}_1^2(Y_i, y)].$$

Discretized conditional raw covariances Z_{iG} (7) and Z_{iL} (8) are $m \times m$ nonnegative definite matrices with $\{Z_{iG}\}_{j,k} = (Y_i(t_j) - \hat{\mu}_G(x, t_j))(Y_i(t_k) - \hat{\mu}_G(x, t_k))$ and $\{Z_{iL,h_1}\}_{j,k} = (Y_i(t_j) - \hat{\mu}_L(x, h_1, t_j))(Y_i(t_k) - \hat{\mu}_L(x, h_1, t_k))$, with $\tilde{d}_2^2(Z_1, Z_2) = \sum_{j=1}^m \sum_{k=1}^m (Z_1(t_j, t_k) - Z_2(t_j, t_k))^2$ denoting the matrix Frobenius norm. The conditional covariance estimates are then

$$\hat{C}_G(x, t_j, t_k) = \underset{\omega \in \tilde{\Omega}}{\arg\min} \, n^{-1} \sum_{i=1}^n w_{iG}(x)\tilde{d}_2^2(Z_{iG}, \omega), \tag{20}$$

$$\hat{C}_{L,h_1,h_2}(x, t_j, t_k) = \underset{\omega \in \tilde{\Omega}}{\arg\min} \, n^{-1} \sum_{i=1}^n w_{iL}(x, h_2)\tilde{d}_2^2(Z_{iL,h_1}, \omega), \tag{21}$$

where $\tilde{\Omega}$ is the space of nonnegative definite matrices on $\mathbb{R}^{m \times m}$.

Setting $\tilde{C}_G(x) = n^{-1} \sum_{i=1}^n w_{iG}(x)Z_{iG}$ and $\tilde{C}_{L,h_1}(x) = n^{-1} \sum_{i=1}^n w_{iL}(x, h_1)Z_{iL,h_1}$ as in (11), with eigenvalue decomposition $\sum_{k=1}^m \tilde{\lambda}_{kG}v_{kG}v_{kG}^T$ and $\sum_{k=1}^m \tilde{\lambda}_{kL}v_{kL}v_{kL}^T$, respectively, the solutions can be written as

$$\hat{C}_G(x, t_j, t_k) = \left\{ \sum_{\tilde{\lambda}_{kG}>0} \tilde{\lambda}_{kG}v_{kG}v_{kG}^T \right\}_{j,k}, \tag{22}$$

$$\hat{C}_{L,h_1,h_2}(x, t_j, t_k) = \left\{ \sum_{\tilde{\lambda}_{kG}>0} \tilde{\lambda}_{kL}v_{kL}v_{kL}^T \right\}_{j,k}. \tag{23}$$

Note that the local versions depend on two bandwidths h_1, h_2 for estimating the conditional mean and conditional covariance. While h_1 can be chosen by cross-validation, the bias in empirical covariances (7), (8) makes cross-validation for choosing h_2 infeasible. To address this problem, we simplify the selection by setting $h_2 = h_1$ for conditional mean and conditional covariance estimation, where we choose h_1 by cross-validation for the conditional mean estimation, minimizing

$$CV_{Mean}(h) = \sum_{i=1}^n d_2^2(\hat{\mu}_L^{(-i)}(X_i, h_1, \cdot), Y_i), \tag{24}$$

where $\hat{\mu}_L^{(-l)}(X_l, h_1, \cdot)$ is the local Fréchet conditional mean estimation from the reduced sample $\{X_i, Y_i\}, i = 1, 2, \dots, l-1, l+1, \dots, n$.

For the simulations, we generated data $(X_i, Y_i), i = 1, \ldots, n$, with functional responses $Y_i \in L^2([0,1])$ and scalar predictors X_i. The conditional distributions $\mathcal{L}(Y|X)$ were constructed as Gaussian processes and the functional responses $Y(t)$ were assumed to be observed on an equidistant grid with $m = 51$ grid points. We investigated sample sizes $n = 50, 100, 200, 300, 400, 500, 1000$ and generated data $(Y_i(t_j), X_i), i = 1, 2, \ldots, n, j = 1, 2, \ldots, 51, (t_1, t_2, \ldots, t_m) = (0, \frac{1}{m-1}, \frac{2}{m-1}, \ldots, 1)$

Specifically, predictors $X_i, i = 1, 2, \ldots, n$, were generated i.i.d. from Unif([0, 1]) and responses $Y_i, i = 1, 2, \ldots, n$, independently from

$$Y(t) = \sqrt{3}\theta_1 t + \sqrt{\frac{6}{5}}\theta_2 \left(1 - \frac{t}{2}\right),$$

with $\theta_1|X \sim N(X, X^2)$, $\theta_2|X \sim N(X/2, (1-X)^2)$, independent given X. For this setting, one can analytically determine the true conditional means and covariances of $\mathcal{L}(Y|X)$,

$$E(Y(t)|X) = \sqrt{3}Xt + \sqrt{\frac{6}{5}}\frac{X}{2}\left(1 - \frac{t}{2}\right),$$
$$\text{Cov}(Y(s), Y(t)|X) = 3X^2 ts + \frac{6}{5}(1-X)^2\left(1 - \frac{t}{2}\right)\left(1 - \frac{s}{2}\right).$$

Local Fréchet regression was used to estimate conditional mean and conditional covariance. For an initial study we evaluated the performance of bandwidth choices with bandwidths h_1 and h_2, selected as combinations of 10 log scale equidistant bandwidths from 0.01 to 0.5. We evaluated the accuracy of the conditional distribution estimation by the average Wasserstein distance (19) over all x and the accuracy of the resulting conditional covariance estimates by the mean integrated square error (MISE),

$$\text{MISE} = E\left(\int_0^1 \int_0^1 \int_0^1 (\hat{C}(x, s, t) - C(x, s, t))^2 \, ds \, dt \, dx\right),$$

which was estimated with $Q = 1000$ Monte Carlo runs for each n by the empirical mean integrated square error (EMISE),

$$\text{EMISE} = \frac{1}{Q}\sum_{q=1}^Q \int_0^1 \int_0^1 \int_0^1 (\hat{C}_q(x, s, t) - C(x, s, t))^2 \, ds \, dt \, dx.$$

The dependency of EMISE on bandwidth choices h_1 and h_2 in Table 1 indicates that although the best bandwidth combination is sometimes a little off the diagonal $h_1 = h_2$, the bandwidths on the diagonal provide similar EMISEs as the best combinations. Thus, cross-validation choices with $h_2 = h_1$, selecting h_1 by cross-validation as described above, worked very well overall. Fitted and true conditional covariance surfaces are illustrated in Figure 1, demonstrating reasonable accuracy and goodness-of-fit. The plot of EMISE versus sample size in Figure 2 agrees with the asymptotic predictions in Proposition 5.

Figure 3 demonstrates the performance of prediction bands for level $\alpha = 0.1$ under the global model. There are 85 lines covered by the band among 100 simulated observations. The performance of the prediction bands was evaluated at $x = 0.1, 0.2, \ldots, 0.9$ and sample sizes $n = 100, 500, 1000$. For each Monte Carlo run, we generated a random sample $(Y_i(t_j), X_i)$, $i = 1, 2, \ldots, n, j = 1, 2, \ldots, 51$ from the global model and constructed nine global prediction bands

TABLE 1 Empirical mean integrated square error of different bandwidth h_1 and h_2 at sample size 100

h_1, h_2	0.01	0.024	0.057	0.136	0.324	0.5
0.01	0.18	0.159	0.139	0.132	0.179	0.182
0.024	0.241	0.105	0.095	0.106	0.163	0.163
0.057	0.131	0.119	0.091	0.084	0.117	0.141
0.136	0.167	0.088	0.072	0.064	0.106	0.128
0.324	0.185	0.094	0.062	0.055	0.074	0.104
0.5	0.186	0.143	0.044	0.046	0.078	0.092
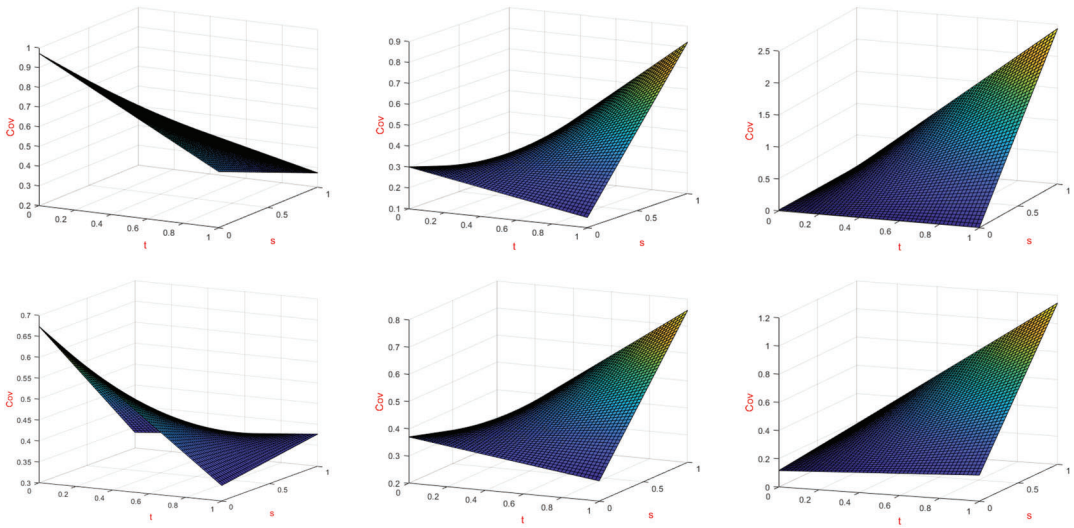
FIGURE 1 True conditional covariances (top 3) and fitted conditional covariances (bottom 3) from one simulation run with sample size $n = 100$. The selected run is of medium empirical mean integrated square error. The left column is when $x = 0.1$, middle column when $x = 0.5$, and right column when $x = 0.9$
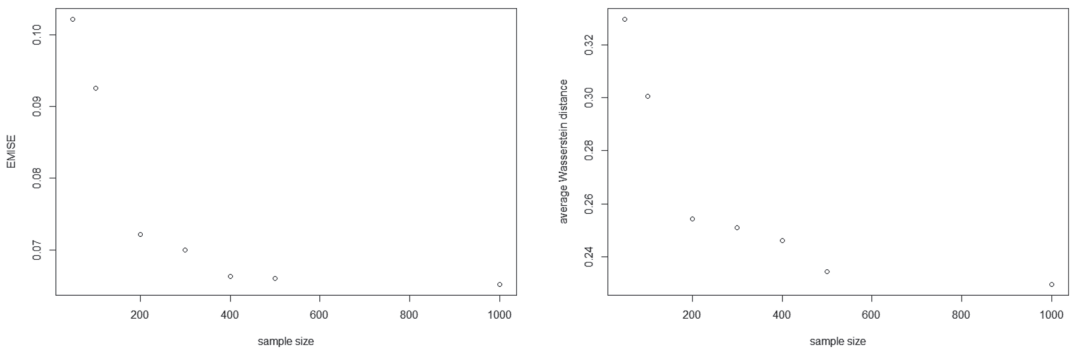
FIGURE 2 EMISE (left) and average Wasserstein distance (right) versus sample size for local Fréchet regression. The bandwidths h are chosen by cross-validation

FIGURE 3 Level 0.9 prediction band (blue solid band) for simulated data at $x = 0.8$ with $n = 100$. Another 100 random observations are generated from the true model to calculated the coverage rate
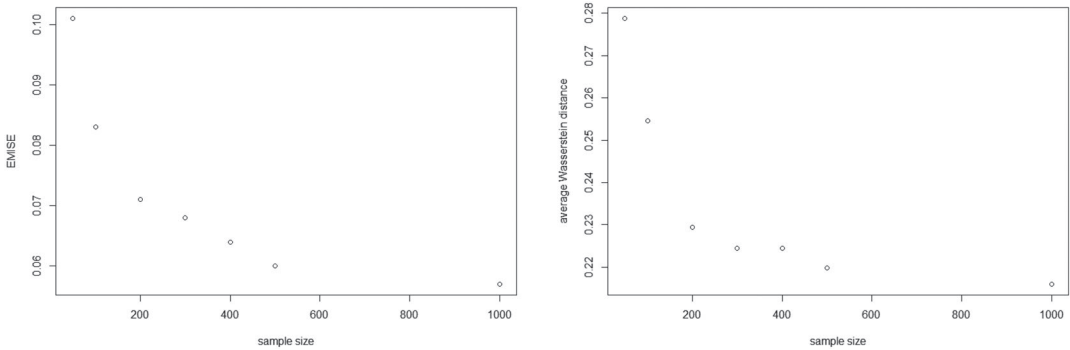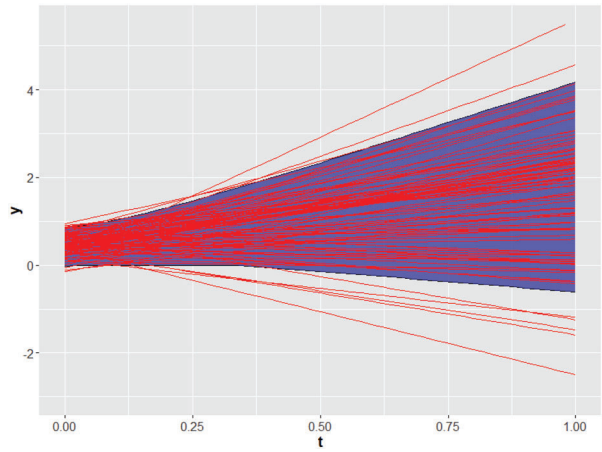

FIGURE 4 Empirical mean integrated square error (left) and average Wasserstein distance (right) versus sample size for global Fréchet regression with 500 Monte Carlo runs

at $x = 0.1, 0.2, \ldots, 0.9$. Then 100 new observations were generated from the true model at each x and the number of curves falling within the constructed prediction bands was recorded. For sample sizes $n = 100, 500, 1000$, we generated 500 Monte Carlo runs and the coverage rate was calculated by averaging over all x and Monte Carlo runs, leading to results of 0.874, 0.888, 0.891 for the respective sample sizes. As sample size gets larger, the coverage of the band is seen to become more accurate.

As mentioned above, global Fréchet regression relies on stronger model assumptions and is only consistent when the true conditional covariance and conditional mean are linear in X as shown above. Therefore, we generated data according to $Y(t) = \theta_1 \sqrt{3}t + \theta_2 \sqrt{(6/5)}(1 - t/2)$, where $\theta_1 | X \sim N(X, X)$, $\theta_2 | X \sim N(X/2, 1 - X)$ and independent given X. Then the true conditional mean and covariance are $E(Y(t)|X) = \sqrt{3}Xt + \frac{X}{2}\sqrt{(6/5)}(1 - t/2)$, $\text{Cov}(Y(s), Y(t)|X) = 3Xts + (6/5)(1 - X)(1 - t/2)(1 - s/2)$. Under this setting we chose different sample sizes and calculated EMISE and average Wasserstein distance as in Figure 4. The curve on the left of Figure 4 is close to $n^{-1/2}$, in accordance with the theoretical result in Proposition 3 and the average Wasserstein distance result is similar to that obtained for the local method in Figure 2.

5 | DATA ILLUSTRATIONS

5.1 | Bike-sharing Data

The bike-sharing data are available at http://capitalbikeshare.com/system-data (Fanaee-T & Gama, 2014). The counts of bike-sharing rentals were recorded every hour in Washington, DC for 2 years, along with weather and seasonal covariates such as temperature, wind speed, and an indicator whether the day of the rental is a holiday. Our goal is to study how the conditional distribution and prediction bands for the bike-sharing rental process change in dependency on these covariates.

With data available for $n = 731$ days, the functional response Y_i is the bike-sharing rental process on the ith day, with data recorded every hour as the number of rentals during the hour (there are a few missings). The covariates are indicators for Spring, Summer, Autumn, work day, clouds, slight rain or snow, heavy rain or snow, and in addition continuous predictors that include average temperature of the day, average humidity of the day (scaled to $[0, 1]$), and average wind speed of the day (also scaled to $[0, 1]$).

Since the dimension of X is high and there are several binary predictors, local Fréchet regression is not feasible, but we can apply global Fréchet regression to obtain conditional distributions of the rental process. The effects of season and work days versus holidays are illustrated in Figure 5. This indicates that weather primarily influences the scale of bike rentals, while whether it is a holiday or work day has a major effect on shape and width of the prediction band. During the work week, peaks in the profile are present at around 8 a.m. and 5 p.m., corresponding to the major commuting times, so clearly on week days one of the main purposes of renting a bike is to use it to commute to the workplace.

By contrast, on holidays, a broad peak in the profile tends to occur around 12–3 p.m. and there are more people renting bikes at midnight than at work days. Unsurprisingly, in Winter fewer people rent a bike than in Spring, Summer, and Autumn. In Summer, the bike rentals at the evening peak have a wider prediction band, indicating more variability, and there is some graphical evidence for an afternoon peak that is much less expressed in the other seasons. In Figure 6 we show the prediction bands for the rental processes in dependence on wind speed. For higher wind speeds, the prediction bands widen and there are overall less bike rentals, as one would expect.

5.2 | Fragments of growth curves

Fragments are a special type of functional data, where for each subject observations on only a part of the time domain are available. Recently, this kind of data, which are also referred to as functional snippets, has become a topic of much research and discussion in functional data analysis due to the ubiquity of functional data fragments in applications (Delaigle & Hall, 2013, 2016; Kraus, 2019; Liebl & Rameseder, 2019; Stefanucci et al., 2018). We provide here some tools for such data that are a natural consequence of the proposed conditional distribution analysis for functional responses. This type of data is for example encountered in accelerated longitudinal studies (Dawson & Müller, 2018; Galbraith et al., 2017). In such studies, the full time domain on which a dynamic process of interest unfolds extends over a longer stretch of time, while only a relatively short observational study can be carried out. Then subjects at different stages of the full longitudinal process are simultaneously observed for a relatively short time interval that depends
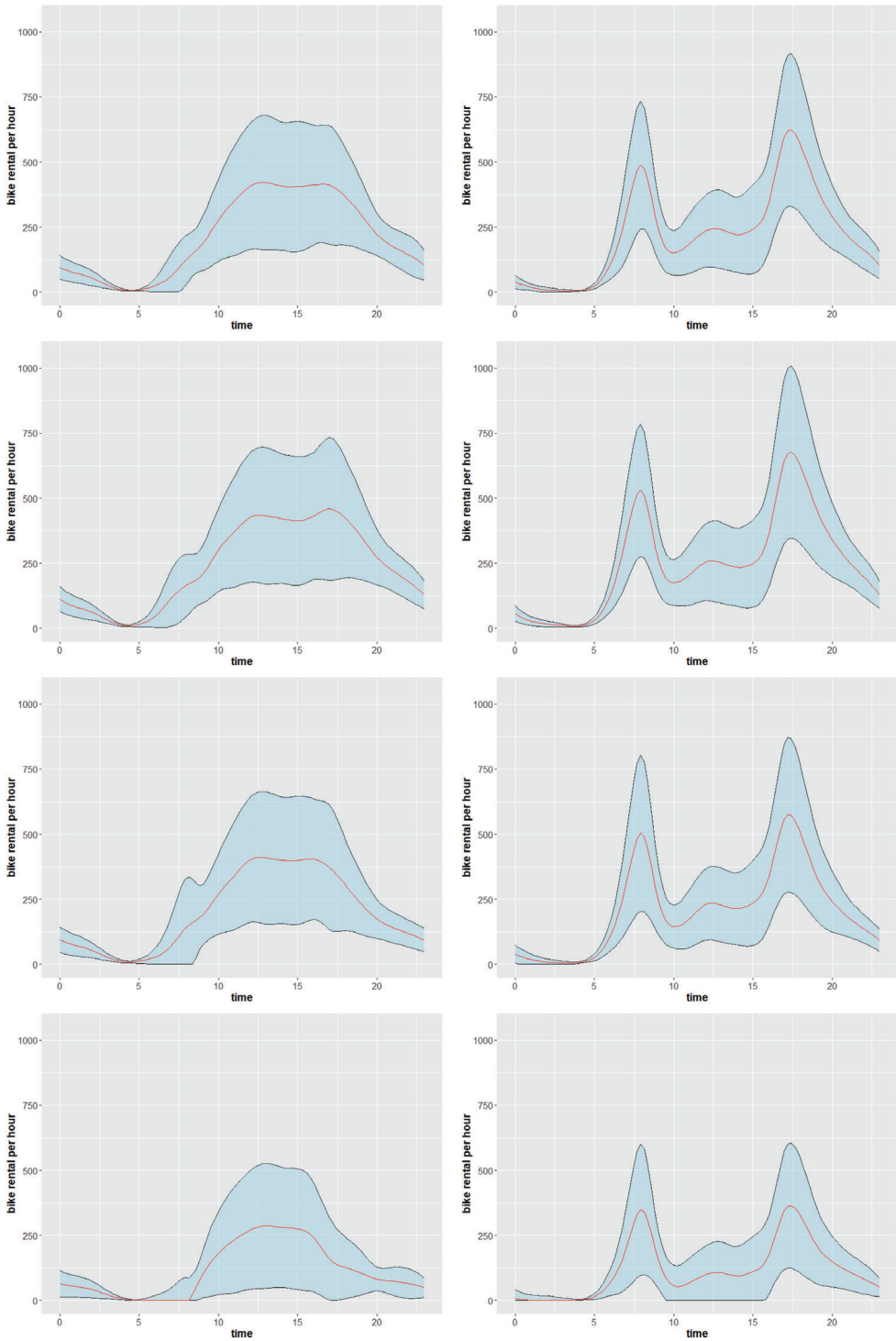
FIGURE 5 Approximate simultaneous prediction bands at level 0.8 for different seasons and weekdays/weekends. Top row conditional on Spring, second row on Summer, third row on Autumn and bottom row on Winter; left column results are conditional on holidays or weekends and right column results conditional on work days. For all cases the wind speed is chosen as the average of the season. The red curves are estimated conditional means
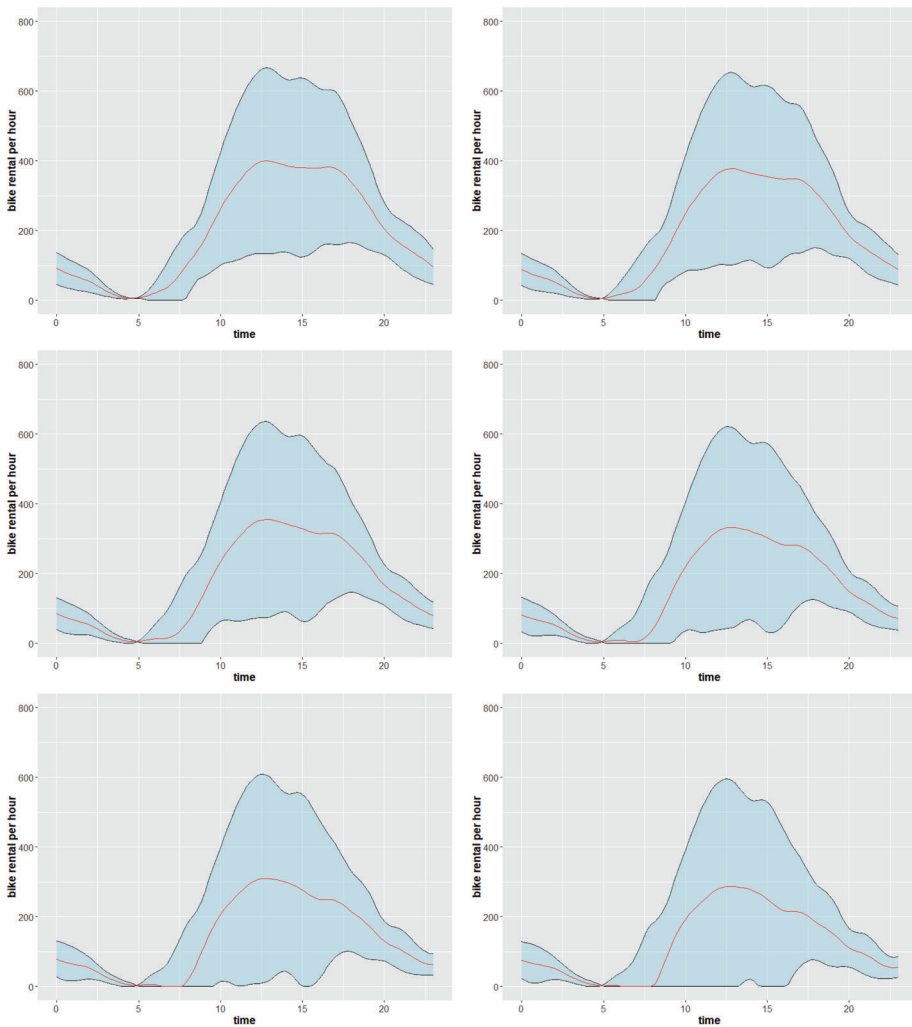
FIGURE 6 Approximate simultaneous prediction bands at level 0.8 conditional on different levels of wind speed on holidays or weekends in Spring. Top panels are with wind speed 0.3 and 0.4 from left to right, the middle panels 0.5 and 0.6, and the bottom panels 0.7 and 0.8. The red curves are estimated conditional means

on the duration of the accelerated longitudinal study, and subjects enter the study at a random time distance from the time origin of the longitudinal process of interest.

Formally, we have a stochastic process $Y(t)$, $t \in [0, T]$, where in the following we refer to the time t as age of the subject. Each subject is observed over an interval of short fixed length $T_0 < T$ and the age s_i when the ith subject enters the study is random and i.i.d. We assume that during the observation period the trajectories of subjects who enter the study are fully observed or that the $Y_i(t)$, $t \in [s_i, s_i + T_0]$ are available on a dense grid. Then the observed data are $\{Y_i(t_{ij}), s_i\}$, $t_{ij} \in [s_i, s_i + T_0]$, $i = 1, 2, \ldots, n$, where t_{ij}, $j = 1, 2, \ldots, m_i$, is the dense time grid where subject i is observed.

We illustrate the application of our methods to this problem by drawing fragment samples from the Zürich longitudinal growth study (Gasser & Kneip, 1995; Kneip & Gasser, 1988).
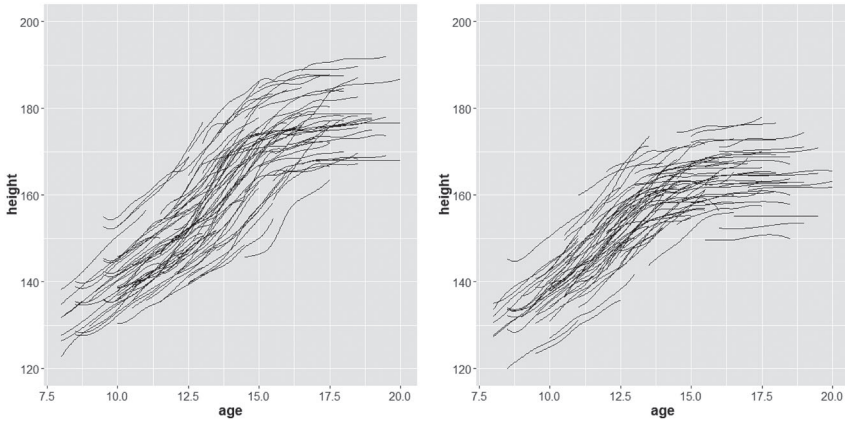
FIGURE 7 Fragment growth curves of $n = 120$ boys and $n = 112$ girls. The left are boys and right are girls
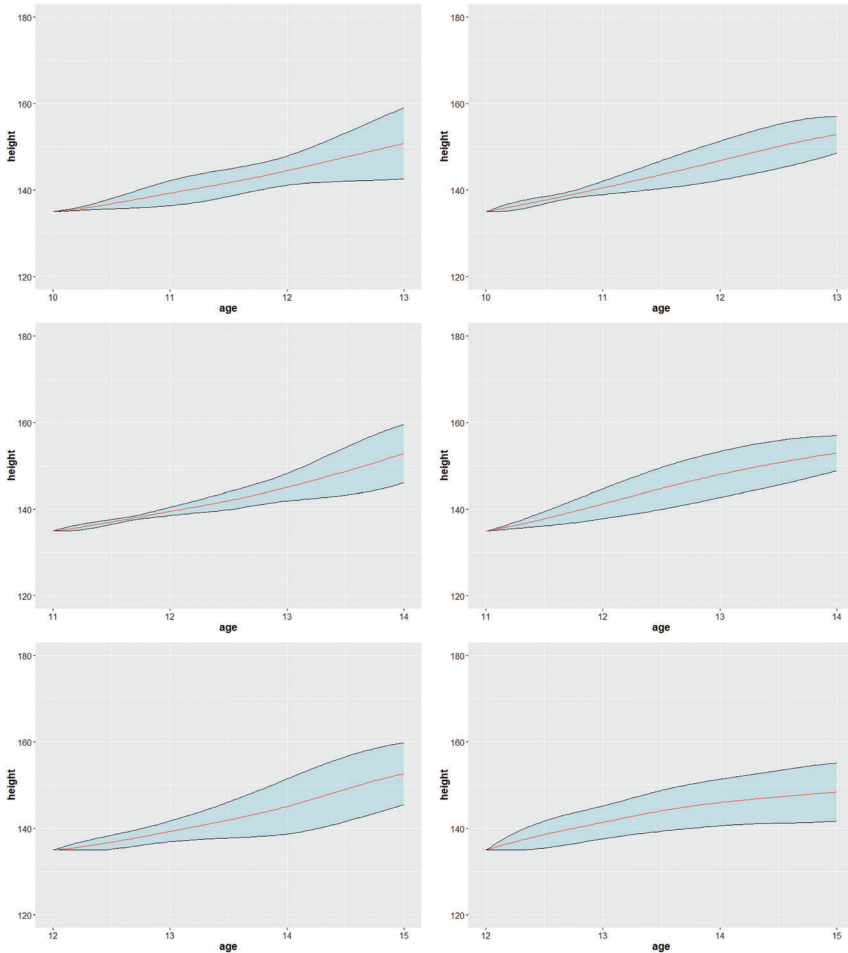

FIGURE 8 Prediction bands at level 0.9 over fragments for children entering at age 10, 11, and 12 years and at height 135 cm. The top panels are with entering age 10, the middle panels with entering age 11, and the bottom panels with entering age 12. The left column corresponds to boys results and the right corresponds to girls. The red curves are estimated conditional means

The Zürich longitudinal growth study includes body heights of 120 boys and 112 girls at different ages from 0 to 20 years. The study consists of longitudinal measurements recorded four times between 0 and 1, half-yearly between 10 and 18, and yearly otherwise with the last measurement at age 20 years. For each subject we assign a random entering age, which is uniformly drawn from a grid of ages that spans half-yearly between 10 and 17. The length of each fragment is chosen to be 3 years, that is, we assume the accelerated longitudinal study is conducted over an interval of 3 years. The fragments drawn are as shown in Figure 7.

We apply local Fréchet regression to obtain $\mathcal{L}(Y(s+t)|s, Y(s))$, $t \in [0, T_0]$, as described in Section 2, aiming to obtain the conditional distribution of the observed trajectories on the fragment in dependence on the value of the trajectory upon entering the fragment and the age at which a subject enters the fragment. These quantities are known at the time when the fragment period starts and the resulting conditional distribution is of interest for various applications, for
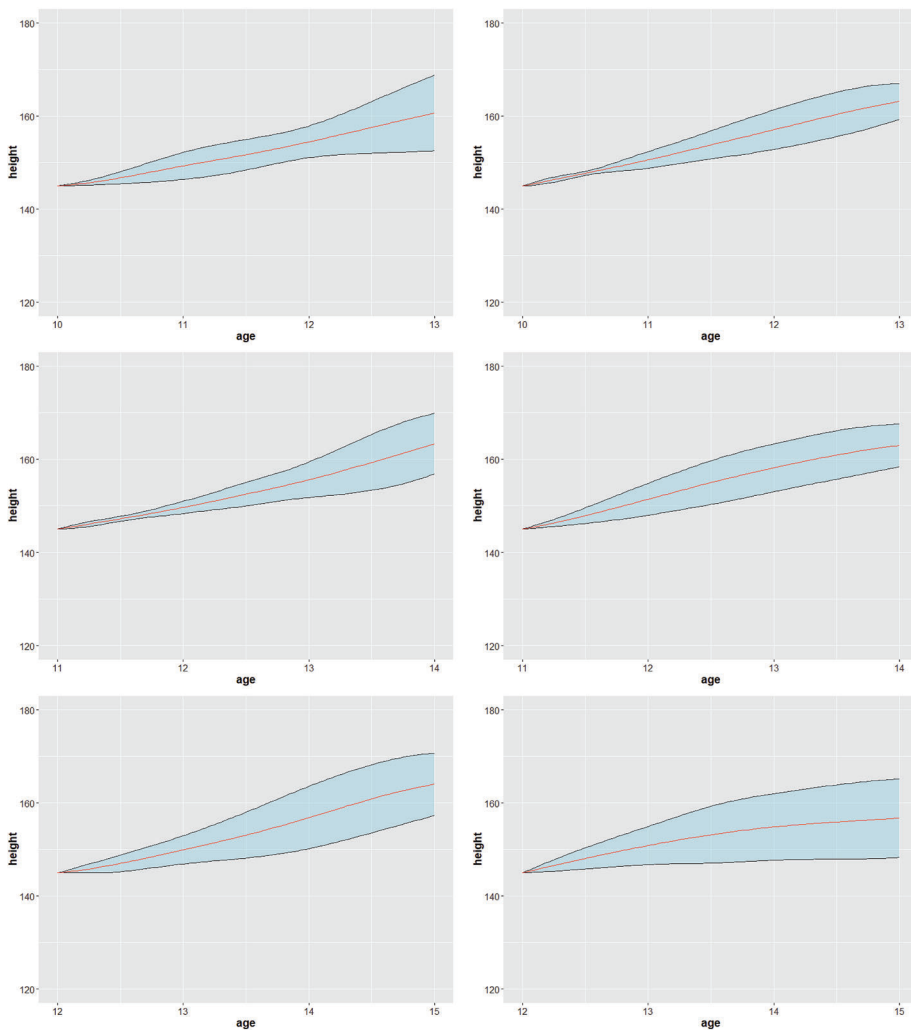

FIGURE 9 Prediction bands at level 0.9 over fragments for children entering at age 10, 11, and 12 years and at height 145 cm. The panels are arranged as in the preceding figure, Figure 8. The red curves are estimated conditional means

example, to determine whether growth is on track during the fragment period. We note here that functional principal component analysis is for example not feasible for functional fragments unless one imposes extremely strong and unverifiable assumptions, and there are very few options available to analyze such data if one wishes to avoid what amounts to essentially parametric assumptions. The method we propose here is one of the available options that can be employed under mild assumptions.

To gain insights into fragmented functional data, we use the conditional simultaneous prediction region (16) that results from conditional distribution modeling. Here we condition on age at entry and value of body length (which generates the functional process) at age at entry into the domain of the fragment. The prediction band in this case is for the interval $[s, s + T_0]$.
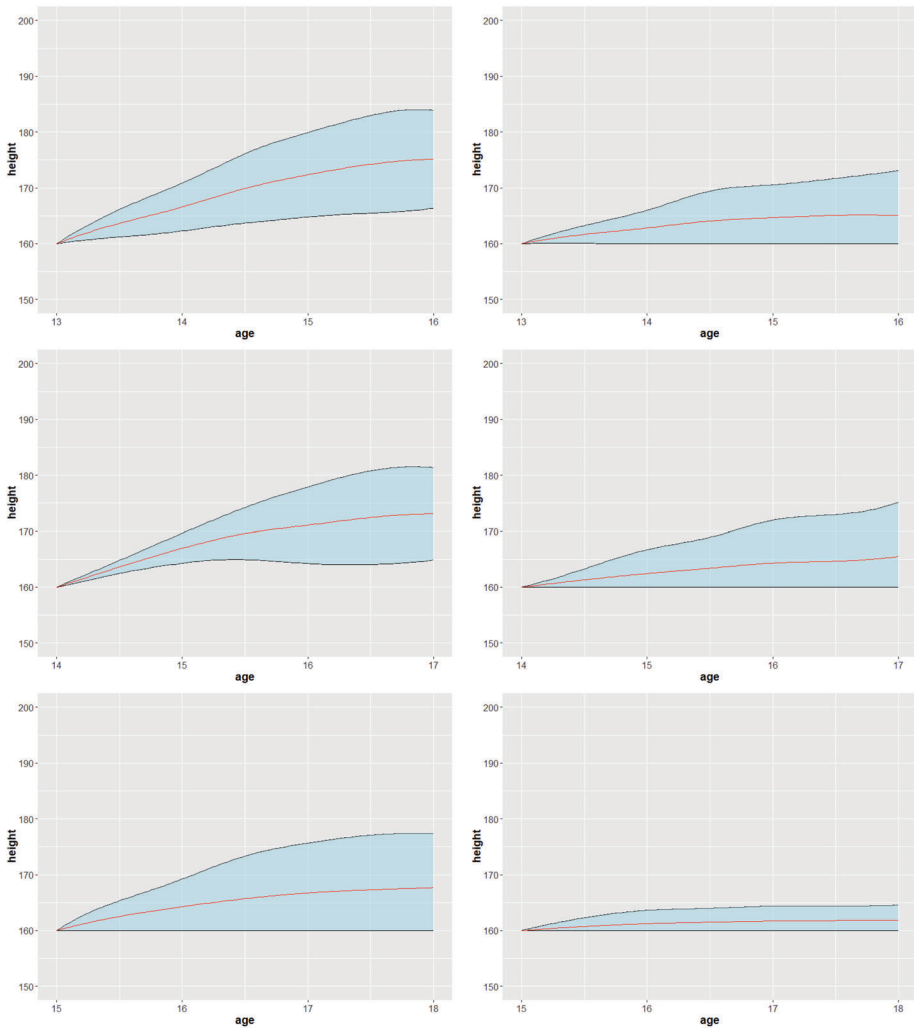

FIGURE 10 Prediction bands at level 0.9 for children entering at age 13, 14, and 15 years and at height 160 cm. The top panels are with entering age 1, middle panel with entering age 14, and bottom panel with entering age 15. The results for boys are in the left column and those for girls in the right column. The red curves are estimated conditional means
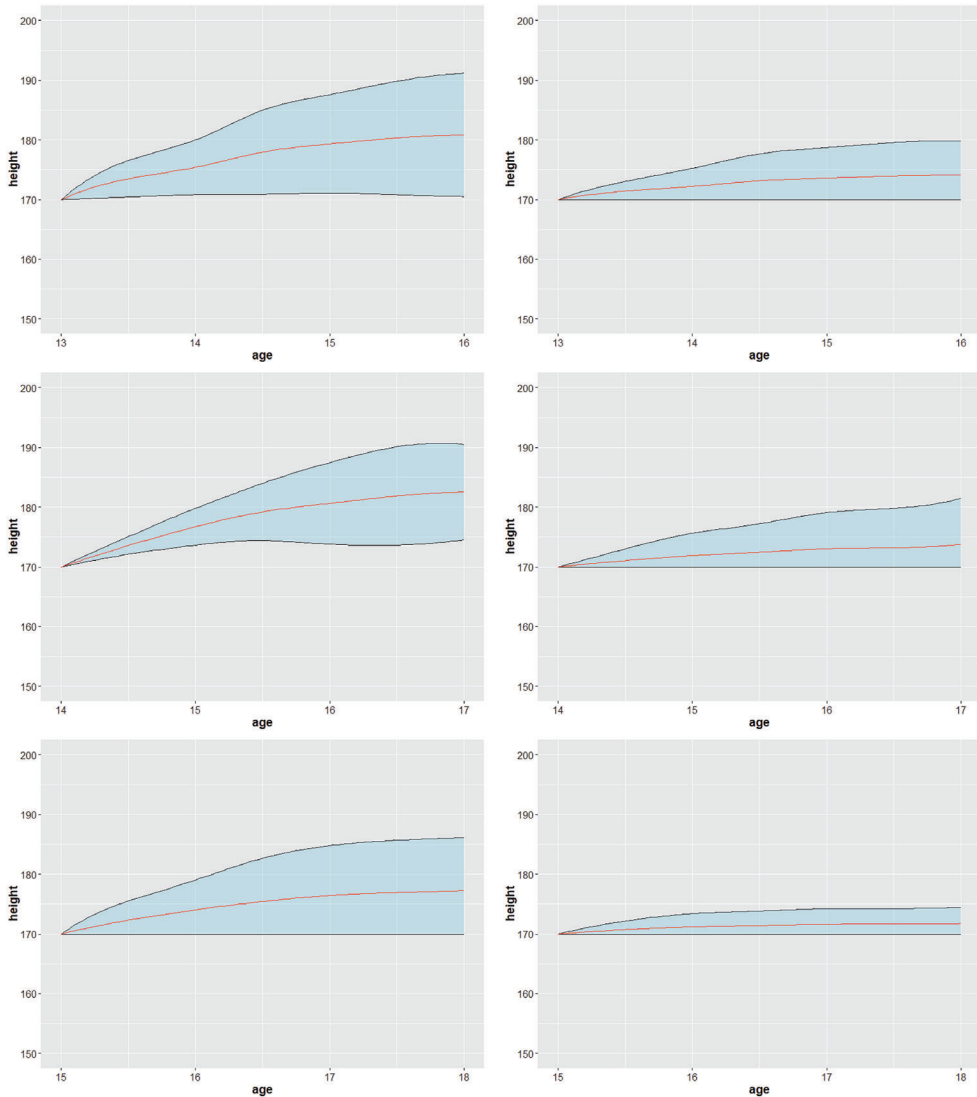
FIGURE 11   Prediction bands at level 0.9 for children entering at age 13, 14, and 15 years and at height 170 cm. The panels are arranged as in Figure 10. The red curves are estimated conditional means

We present the results of prediction bands in Figures 8, 9, 10, and 11 for entering ages 10, 11, and 12 years, and entering heights 135 and 145 cm, and also conditioning on entering ages 13, 14, and 15 years and on entering heights of 160 and 170 cm. At entering age 10, 11, and 12 years and same entering height, boys and girls have similar conditional mean and variation while at entering ages 13, 14, and 15 years, the conditional means of boys are much higher and the prediction bands for boys are much wider. For both boys and girls at the same entering height, larger age at entry into the fragment domain is associated with similar conditional means but higher variation at age 10, 11, and 12 years; however, at entering ages 13, 14, and 15 years, larger entering ages are associated with less variation, which can be explained by the fact that growth ends around age 18 years.

6 | DISCUSSION

In the article, we propose both linear and nonparametric models to construct conditional distributions for regression models with functional responses, thus broadening the scope of functional regression, which is normally confined to the usual mean regression. This approach can also be harnessed to construct asymptotic simultaneous prediction bands, which can be used to determine whether new trajectory observations that are associated with certain covariate levels are outliers or fall into the expected range. The proposed methods are supported by asymptotic consistency, including Wasserstein consistency for conditional distributions.

The proposed method is shown to be also useful for longitudinal fragment data that are commonly encountered in accelerated longitudinal studies. Approaches that aim to estimate the whole covariance surface are not useful for functional data that consist of genuine fragments. Instead, we solve the conditional distribution problem under Gaussian assumptions, which can then be harnessed to assess the dynamic behavior of subjects within the fragments conditional on the age at entry, as well as the function value at the beginning of the fragment, where the distribution of the functional response during the fragment is predicted from the data available at entry into the fragment.

ORCID
Jianing Fan https://orcid.org/0000-0002-1277-9096

REFERENCES

Bosq, D. (2012). *Linear processes in function spaces: Theory and applications*. Springer.

Cardot, H. (2007). Conditional functional principal components analysis. *Scandinavian Journal of Statistics*, *34*, 317–335.

Chen, K., & Müller, H.-G. (2012). Conditional quantile analysis when covariates are functions, with application to growth data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *74*, 67–89.

Chen, K., & Müller, H.-G. (2014). Modeling conditional distributions for functional responses, with application to traffic monitoring via GPS-enabled mobile phones. *Technometrics*, *56*, 347–358.

Chiou, J.-M., & Müller, H.-G. (2009). Modeling hazard rates as functional data for the analysis of cohort lifetables and mortality forecasting. *Journal of the American Statistical Association*, *104*, 572–585.

Chiou, J.-M., Müller, H.-G., & Wang, J.-L. (2003). Functional quasi-likelihood regression models with smooth random effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*, 405–423.

Chiou, J.-M., Müller, H.-G., & Wang, J.-L. (2004). Functional response models. *Statistica Sinica*, *14*, 675–693.

Chowdhury, J., & Chaudhuri, P. (2016). Nonparametric depth and quantile regression for functional data. *arXiv preprint arXiv:1607.03752*.

Cuesta-Albertos, J., Matrán-Bea, C., & Tuero-Diaz, A. (1996). On lower bounds for the L2-Wasserstein metric in a Hilbert space. *Journal of Theoretical Probability*, *9*, 263–283.

Dawson, M., & Müller, H.-G. (2018). Dynamic modeling of conditional quantile trajectories, with application to longitudinal snippet data. *Journal of the American Statistical Association*, *113*, 1612–1624.

Delaigle, A., & Hall, P. (2013). Classification using censored functional data. *Journal of the American Statistical Association*, *108*, 1269–1283.

Delaigle, A., & Hall, P. (2016). Approximating fragmented functional data by segments of Markov chains. *Biometrika*, *103*, 779–799.

Ding, H., Zhang, R., & Zhang, J. (2018). Quantile estimation for a hybrid model of functional and varying coefficient regressions. *Journal of Statistical Planning and Inference*, *196*, 1–18.

Fanaee-T, H., & Gama, J. (2014). Event labeling combining ensemble detectors and background knowledge. *Progress in Artificial Intelligence*, *2*, 113–127.

Faraway, J. J. (1997). Regression analysis for a functional response. *Technometrics*, *39*, 254–261.

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré*, *10*, 215–310.

Galbraith, S., Bowden, J., & Mander, A. (2017). Accelerated longitudinal designs: An overview of modelling, power, costs and handling missing data. *Statistical Methods in Medical Research*, *26*, 374–398.

Gasser, T., & Kneip, A. (1995). Searching for structure in curve samples. *Journal of the American Statistical Association*, *90*, 1179–1188.

Jiang, C.-R., & Wang, J.-L. (2010). Covariate adjusted functional principal components analysis for longitudinal data. *The Annals of Statistics*, *38*, 1194–1226.

Kato, K. (2012). Estimation in functional linear quantile regression. *The Annals of Statistics*, *40*, 3108–3136.

Kneip, A., & Gasser, T. (1988). Convergence and consistency results for self-modeling nonlinear regression. *Annals of Statistics*, *16*, 82–112.

Koenker, R., & Bassett, G., Jr. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, *46*, 33–50.

Kraus, D. (2019). Inferential procedures for partially observed functional data. *Journal of Multivariate Analysis*, *173*, 583–603.

Liebl, D., & Rameseder, S. (2019). Partially observed functional data: The case of systematically missing parts. *Computational Statistics & Data Analysis*, *131*, 104–115.

Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application*, *2*, 321–359.

Petersen, A., Deoni, S., & Müller, H.-G. (2019). Fréchet estimation of time-varying covariance matrices from sparse data, with application to the regional co-evolution of myelination in the developing brain. *The Annals of Applied Statistics*, *13*, 393–419.

Petersen, A., & Müller, H.-G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *Annals of Statistics*, *44*, 183–218.

Petersen, A., & Müller, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Annals of Statistics*, *47*, 691–719.

Ramsay, J. O. (2006). *Functional data analysis*. Wiley.

Ramsay, J. O., & Dalzell, C. J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B*, *53*, 539–572.

Stefanucci, M., Sangalli, L. M., & Brutti, P. (2018). PCA-based discrimination of partially observed functional data, with an application to AneuRisk65 data set. *Statistica Neerlandica*, *72*, 246–264.

Takatsu, A. (2011). Wasserstein geometry of Gaussian measures. *Osaka Journal of Mathematics*, *48*, 1005–1026.

van der Vaart, A., & Wellner, J. (1996). *Weak convergence and empirical processes with applications to statistics*. Springer.

Villani, C. (2008). *Optimal transport: Old and new*. Springer.

Wang, H. J., Zhu, Z., & Zhou, J. (2009). Quantile regression in partially linear varying coefficient models. *The Annals of Statistics*, *37*, 3841–3866.

Wang, J.-L., Chiou, J.-M., & Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, *3*, 257–295.

Yao, F., Sue-Chee, S., & Wang, F. (2017). Regularized partially functional quantile regression. *Journal of Multivariate Analysis*, *156*, 39–56.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.