Nonparametric regression method with functional covariates and multivariate response

Kurdistan M.Taher Omar[1*]   Bo Wang[2]

[1]Department of Mathematics, University of Leicester, Leicester LE1 7RH, UK; and Department of Mathematics, Faculty of Science, University of Zakho, Kurdistan Region, Iraq
[2]Department of Mathematics, University of Leicester, Leicester LE1 7RH, UK

November 18, 2017

Abstract

Nonparametric regression methods have been widely studied in functional regression analysis in the context of functional covariates and univariate response, but it is not the case for functional covariates with multivariate response. In this paper, we present two new solutions for the latter problem: the first is to directly extend the nonparametric method for univariate response to multivariate response, and in the second, the correlation among different responses is incorporated into the model. The asymptotic properties of the estimators are studied, and the effectiveness of the proposed methods is demonstrated through several simulation studies and a real data example.

Keywords: Functional data analysis; Functional regression; Functional covariates; Multivariate response; Semi-metrics; Almost complete convergence

1   Introduction

We start with a motivating example which concerns a dataset from a quality control problem in the food industry. The data are recorded on a Tecator Infratec Food and Feed Analyzer working in the wavelength range 850 - 1050nm by the Near Infrared Transmission (NIT) principle. Each sample contains one of 215 finely chopped pure meat with different moisture (water), fat and protein contents. For each meat sample the data consists of a 100 channel spectrum of absorbances and the contents of water, fat and protein. The spectrometric curves are shown in Fig. 1. The task is to predict the three contents from the spectrometric curves.

---

[*]Corresponding author. Email: kmtto2@le.ac.uk and kurdistanmt@gmail.com
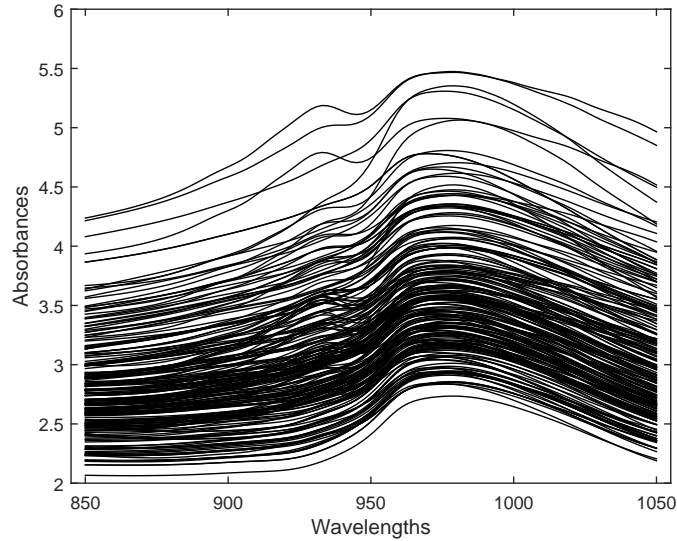
Figure 1: Spectrometric curves.

The above is a typical problem in functional data analysis. Functional data concerns data which are collected as curves, surfaces or measurements varying over a continuum. With the fast advances of technologies functional data have become more and more prevalent in a large number of fields, such as environment, medicine, finance, industry and social science. Consequently functional data analysis has received increasing attention, particularly after the work of Ramsay and Silverman (1997, 2005) was published.

For functional data analysis two main streams of methodogies exist in the literature: functional parametric methods and functional nonparametric methods. Ramsay and Silverman (1997, 2005) studies in details a number of functional parametric models, which have been further extended and developed by many researchers. Ferraty and Vieu (2003, 2004, 2006) first introduces nonparametric methods for functional data analysis with functional covariates and univariate scalar response. It is followed by a number of further developments, such as the functional Nadayara-Watson (NW) estimator (Ferraty et al, 2007), the functional k-nearest neighbour estimator (Burba et al, 2009), the functional local linear estimator (Barrientos-Marin et al, 2010), the distance-based local linear estimator (Boj et al, 2010).

Nonparametric modelling has also been proposed for other types of functional data analysis. For instance, Matsui et al. (2008) illustrates the association between multiple scalar responses and functional predictors by using the Gaussian basis function. The problem with multivariate response variables and functional covariates based on the L_1-median regression estimation is described by Chaouch and Laïb (2013). Wang and Chen (2015) proposes the formulation of the covariance function for the multi-response Gaussian process regression. Xiang et al. (2013) studies the multivariate nonparametric regression analysis in the context of longitudinal data. Ferraty et al. (2011, 2012) and Lian (2012) consider the problem of functional regression where both the responses and the explanatory variables are functions.

In this paper, we develop two new solutions for functional regression problem with functional covariates and multivariate response variable. The first solution is to directly extend the nonparametric method for univariate response to multivariate response with different bandwidths for different responses. In the second solution, the correlation among different responses is taken into account and incorporated into the model. Our methods utilize the kernel function with an automatic bandwidth selection by cross-validation procedure and semi-metrics as a measure of the proximity between functional data. The rate of almost complete convergence of the methods is studied. And the effectiveness of the proposed methods is demonstrated through several simulation studies and a real data example.

The rest of the paper is organised as follows. Section 2 presents the model and the estimators. In Section 3, some theoretical properties such as the rate of almost complete convergence are discussed. In Section 4, the effectiveness of the proposed methods is illustrated through several simulation studies and a real data example. Section 5 concludes the paper.

2   Model and Estimation

Let $(\mathcal{X}_i, Y_i)_{i=1,...,n}$ be n pairs of samples independently and identically distributed as $(\mathcal{X}, Y)$ and valued in $f \times R^q$, where f is an infinite-dimensional space equipped with a semi-metric $d(\cdot, \cdot)$. $Y_i = (y_{i1}, ..., y_{iq})^t$ is the q-dimensional response variable and $\mathcal{X}_i$ is the functional predictor. The general framework of the functional regression can be defined by:

$$Y_i = r(\mathcal{X}_i) + \varepsilon_i, \quad i = 1, ..., n, \tag{1}$$

where $r(\cdot)$ is an unknown nonlinear operator, and ε_i is independent random errors with $E[\varepsilon_i|\mathcal{X}_i] = 0$. Therefore, the objective is to estimate the nonlinear operator $r(\chi) = E[Y|\mathcal{X} = \chi]$ for a given value $\mathcal{X} = \chi \in f$. By the nonparametric kernel method, the regression estimator can be expressed as:

$$\hat{r}(\chi) = \sum_{i=1}^{n} \omega_i(\chi) Y_i, \tag{2}$$

where $\omega_i(\chi)$ is the weights.

In this paper, we propose two methods for the estimation of the weights $\omega_i(\chi)$. Firstly, we extend the method for univariate scalar response case to the multivariate case but using different bandwidths for different components of the response. This method is denoted by MRD and given as follows:

$$\omega_i(\chi) = (\sum_{i=1}^{n} \breve{\omega}_i)^{-1/2} \breve{\omega}_i (\sum_{i=1}^{n} \breve{\omega}_i)^{-1/2},$$

where

$$\breve{\omega}_i = \begin{bmatrix} a_{11}^i & 0 & 0 & \dots & 0 \\ 0 & a_{22}^i & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{qq}^i \end{bmatrix},$$

and

$$a_{11}^i = K\left(h_1^{-1} d\left(\chi, \mathcal{X}_i\right)\right)$$

$$a_{22}^i = K\left(h_2^{-1} d\left(\chi, \mathcal{X}_i\right)\right)$$

$$\vdots$$

$$a_{qq}^i = K\left(h_q^{-1} d\left(\chi, \mathcal{X}_i\right)\right).$$

Here $K(\cdot)$ is the kernel function, and $h_1, h_2, ..., h_q$ are the bandwidths for different components of the response.

In the second method (denoted by MRC), we take the correlations between different components of the response variable into account and use different bandwidths for different components.

Since $(\mathcal{X}_i, Y_i)$, $i = 1, ..., n$, are independent and identically distributed, we can define

$$V = \begin{bmatrix} V_{11} & V_{12} & \dots & V_{1q} \\ V_{21} & V_{22} & \dots & V_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ V_{q1} & V_{q2} & \dots & V_{qq} \end{bmatrix} = Cov(Y_i|\mathcal{X}_i).$$

3

Therefore, V is the $q \times q$ covariance matrix of Y_i, where the diagonal elements give the variances of each component of the response and the off-diagonal elements give the covariances among different components.

Now let $\widehat{V}$ be an estimate of V, i.e. the sample covariance matrix of Y_i, and define

$$\mathbf{K}_{iH} = diag\{K(h_l^{-1}d(\mathcal{X}_i, \chi)), \quad l = 1, 2, ..., q\},$$

and

$$\hat{\omega}_i = (\mathbf{K}_{iH}^{-1/2}\widehat{V}\mathbf{K}_{iH}^{-1/2})^{-1}.$$

Then in the second method the weights $\omega_i(\chi)$ are given by

$$\omega_i(\chi) = (\sum_{i=1}^{n} W_i)^{-1/2}\hat{\omega}_i(\sum_{i=1}^{n} W_i)^{-1/2},$$

where W_i is a diagonal matrix containing the row sums of the matrix $\hat{\omega}_i$.

It is noted that we need a q-dimensional bandwidth vector $H = (h_1, \ldots, h_q)$ to allow different degrees of smoothing in different components of the response. In practice, the bandwidth is critical to achieving accurate results. Ferraty and Vieu (2004, 2006), Ferraty et al. (2007) and Burba et al. (2009) use the cross-validation procedure for selecting bandwidths. Similarly, we can also select bandwidth by cross-validation procedure. Let

$$CV_l(h_l) = \frac{1}{n}\sum_{i=1}^{n}(y_{il} - \hat{r}_{l,-i}(\mathcal{X}_i))^2, \quad l = 1, ..., q$$

and

$$CV(H) = CV_1(h_1) + CV_2(h_2) + ... + CV_q(h_q),$$

where $\hat{r}_{l,-i}(\cdot)$ is the estimate of $r_l(\cdot)$ without the ith sample. Therefore, the optimal bandwidths can be determined by minimising $CV(H)$.

3 Asymptotic properties

Let x be a given point in f, and denote by $B(x, h)$ the closed ball of centre x and radius h, namely:

$$B(x, h) = \{x' \in f : d(x, x') \leqslant h\}.$$

The model requires that the probability of $\mathcal{X}$ is such that there exists a non-decreasing function ϕ_x such that:

(H1) $\exists(C_1, C_2), \forall x \in f, \forall \varepsilon > 0,$

$$0 < C_1\phi_x(\varepsilon) \leqslant P(\mathcal{X} \in B(x, \varepsilon)) \leqslant C_2\phi_x(\varepsilon) < \infty.$$

And the joint distribution of $(\mathcal{X}, Y)$ needs to satisfy:

(H2) $\exists C_3, \forall r > 1, E(\|Y\|^r|\mathcal{X}) < C_3r! < \infty.$

(H3) $\exists C_4, \exists b > 0, \exists \gamma > 0, \forall x, x' \in f,$
$$\|r(x) - r(x')\| \leqslant C_4d^b(x, x').$$

In addition, we also need the following technical conditions on the kernel function and the bandwidth.

(H4) The kernel function has to be such that:

   (i) K is a bounded and Lipschitz continuous function with support $[0, 1)$, and if $K(1) = 0$ it has to fulfill, together with $\phi_x(\cdot)$, the conditions:

(ii) $\exists (C_5, C_6) > 0$, such that $-\infty < C_5 \leqslant K'_{iH} \leqslant C_6 < 0$.

(iii) $\exists C_7 > 0, \exists \gamma_0 > 0, \forall \gamma < \gamma_0$,

$$\int_0^\gamma \phi_x(u) du > C_7 \gamma \phi_x(\gamma).$$

(H5) The bandwidth h is a positive sequence such that:

$$\lim_{n \to \infty} h = 0 \quad and \quad \lim_{n \to \infty} \frac{\log n}{n \phi_x(h)} = 0.$$

It is noted that this set of assumptions are different from the case of functional covariate with univariate response and when both explanatory and response variables are functional, because of the different bandwidths for the different components of the response variable.

The rate of convergence of $\hat{r}(x)$ is stated in Theorem 3.1. As mentioned before, this is the first outcome of this type in a nonparametric functional regression setting when the covariate is functional while the response is multivariate with different bandwidths for each component.

Result B1: Define

$$\mathbf{K}_{iH} = diag\{K(h_l^{-1} d(x_i, x)), \quad l = 1, 2, ..., q\},$$

then, $\exists (C_8, C_9) > 0$ such that

$$C_8 \phi_x(h) \leqslant E[\|\mathbf{K}_{iH}\|] \leqslant C_9 \phi_x(h).$$

This result is apparent when $K(1) > 0$ and can be extended to the continuous kernel, satisfying $H4$ as shown in Lemma 4.4 in Ferraty and Vieu (2006).

Theorem 3.1 *Under the hypotheses $H1 - H5$, we have:*

$$\|\hat{r}(x) - r(x)\| = O(h^b) + O_{a.co.}\left(\sqrt{\frac{\log n}{n \phi_x(h)}} \right).$$

Proof For $i = 1, .., n$, we define

$$\Delta_i = \frac{(\mathbf{K}_{iH}^{-1/2} \widehat{V} \mathbf{K}_{iH}^{-1/2})^{-1}}{E\|(\mathbf{K}_{1H}^{-1/2} \widehat{V} \mathbf{K}_{1H}^{-1/2})^{-1}\|}.$$

Note that $H1$, $H4$ and Result $B1$, ensure that $E\|(\mathbf{K}_{1H}^{-1/2} \widehat{V} \mathbf{K}_{1H}^{-1/2})^{-1}\| > 0$.

Let $\hat{r}_1(x)$ and $\hat{r}_2(x)$ be the following quantities:

$$\hat{r}_1(x) = \frac{1}{n} \sum_{i=1}^n \Delta_i$$

and

$$\hat{r}_2(x) = \frac{1}{n} \sum_{i=1}^n Y_i \Delta_i.$$

Then $\hat{r}(x) = \hat{r}_2(x) \hat{r}_1^{-1}(x)$. Our proof depends on the decomposition:

$$\hat{r}(x) - r(x) = \left\{ (\hat{r}_2(x) - E\hat{r}_2(x)) - (r(x) - E\hat{r}_2(x)) \right\} \hat{r}_1^{-1}(x)$$
$$- r(x) \{ \hat{r}_1(x) - 1 \} \hat{r}_1^{-1}(x).$$

Theorem 3.1 will be true providing both the following Lemma 3.2 and Lemma 3.3 can be proved. However, the second part of Lemma 3.3 is addressed directly by using part (i) of Lemma 3.3 when $Y_i = 1$ in combination with part (i) of Proposition A.6 in Ferraty and Vieu (2006).

Lemma 3.2 *Under* $H3, H4,$ *and* $H5$ *we have:*

$$\|r(x) - E\hat{r}_2(x)\| = O(h^b).$$

Proof Model (1) allows us to directly write:

$$
\begin{aligned}
\|r(x) - E\hat{r}_2(x)\| &= \|r(x) - E(Y_1 \Delta_1)\| \\
&= \|r(x) - E[E(Y_1 \Delta_1 \mid X_1)]\| \\
&= \|r(x) - E[E(Y_1 \mid X_1)\Delta_1]\| \\
&= \|r(x) - E(r(X_1)\Delta_1)\| \\
&\leqslant E[\|r(x) - r(X_1)\|\|\Delta_1\|],
\end{aligned}
$$

and by using $H3$, this becomes:

$$\|r(x) - E\hat{r}_2(x)\| \leqslant E[d^b(x, X_1)\|\Delta_1\|].$$

Therefore, with the hypothesis $H1$ and Result $B1$, and since $\|\Delta_1\| = 1, \exists C > 0,$ such that

$$\|r(x) - E\hat{r}_2(x)\| \leqslant Ch^b.$$

The above inequality yields the proof, since C does not depend on x.

Lemma 3.3 *We have:*

(i) Under assumptions $H1 - H5$ *we have:*

$$\|\hat{r}_2(x) - E\hat{r}_2(x)\| = O_{a.co.}\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right).$$

(ii) Under assumptions $H1, H4,$ *and* $H5,$ *we have:*

$$\|\hat{r}_1(x) - 1\| = O_{a.co.}\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right).$$

Proof The idea of this proof is based on using a Bernstein-type exponential inequality. In fact:

$$P\Big(\|\hat{r}_2(x) - E\hat{r}_2(x)\| > \epsilon\Big) = P\Big(\frac{1}{n}\|\sum_{i=1}^{n}(Y_i\Delta_i - E(Y_i\Delta_i)\| > \epsilon\Big),$$

and we have to offer that it exists $\epsilon_0 > 0$ such that:

$$\sum_{n \in N^*} P\Big(\frac{1}{n}\|\sum_{i=1}^{n}(Y_i\Delta_i - E(Y_i\Delta_i)\| > \epsilon_0\sqrt{\frac{\log n}{n\phi_x(h)}}\Big) < \infty.$$

Therefore, applying Corollary $A.8 - ii$ in Ferraty and Vieu (2006) gives the exponential inequality with $Z_i = Y_i\Delta_i - EY_1\Delta_1$. We first need to show that:

$$\exists C_{10} > 0, \quad \forall m = 2, 3, ..., \quad \|E(Y_1\Delta_1 - EY_1\Delta_1)\|^m \leq C_{10}\phi_x(h)^{-m+1}. \tag{3}$$

- We first prove that for $m \geq 2$:

$$E\|Y_1\Delta_1\|^m = O(\phi_x(h)^{-m+1}). \tag{4}$$

For this, we write :

$$E\|Y_1\Delta_1\|^m \leq E[E[\|Y_1\|^m|X_1]\|\Delta_1\|^m].$$

Clearly, we get from the assumption $H2$ that:

$$E\|Y_1\|^m = E[E[\|Y_1\|^m|X]] < Cm! < \infty,$$

which implies that:

$$E\|Y_1\Delta_1\|^m \leq E[E[\|Y_1\|^m|X_1]\|\Delta_1\|^m]$$
$$\leqslant Cm!E\|\Delta_1\|^m.$$

By applying $H4$ and Result B1, we get:

$$\frac{C_5}{\phi_x(h)^{m-1}} \leqslant E\|\Delta_1\|^m \leqslant \frac{C_6}{\phi_x(h)^{m-1}},$$

then, we get:

$$E\|Y_1\Delta_1\|^m = O(\phi_x(h)^{-m+1}). \tag{5}$$

- Furthermore, we utilize the Newton's binomial expansion and get:

$$\|Y_1\Delta_1 - E(Y_1\Delta_1)\|^m = \sum_{k=0}^{m} C_{k,m}\|Y_1\Delta_1\|^k\|E[Y_1\Delta_1]\|^{m-k}(-1)^{m-k},$$

where

$$C_{k,m} = \frac{m!}{k!(m-k)!},$$

which implies that

$$E\|Y_1\Delta_1 - E(Y_1\Delta_1)\|^m \leqslant C\sum_{k=0}^{m} C_{k,m}E\|Y_1\Delta_1\|^k\|r(x)\|^{m-k}$$
$$\leq C \max_{k=0,1,2,...,m} k!\phi_x(h)^{1-k},$$

where C is a real positive constant. Because $\phi_x(h)$ tends to zero when n goes to infinity, it becomes the case that:

$$E\|Y_1\Delta_1 - E(Y_1\Delta_1)\|^m = O((\phi_x(h))^{1-m}).$$

Therefore, we can apply Corollary $A.8 - ii$ in Ferraty and Vieu (2006) with $a^2 = \phi_x(h)^{-1}$.

Subsequently, we have $u_n = (a^2 \log n / n) = \log n / (n\phi_x(h)) \to 0$ as $n \to \infty$, by using assumptions $H4$ and $H5$.

And the second part of the theorem is directly derived from part one by taking $Y_i = 1$, and we get the result:

$$\|\hat{r}_1(x) - 1\| = O_{a.co.}\left(\sqrt{\frac{\log x}{n\phi_x(h)}}\right).$$

Theorem 3.1 shows that the estimator $\hat{r}(x)$ is consistent in terms of almost complete convergence. It also states the rate of pointwise almost complete convergence. As discussed in Ferraty and Vieu (2006), this mode of convergence is stronger than almost sure convergence and convergence in probability, and the links of almost complete convergence with other modes of convergence are also shown in the above reference.

4 Numerical examples

The main aim of this section is to demonstrate the usefulness of the methodologies through simulated and real data examples.

4.1 Simulation studies

Example 1. We first consider the following regression model:

$$Y_i = r(\mathcal{X}_i) + \varepsilon_i, \quad i = 1, 2, ..., n = 215.$$

The functional predictor is generated by

$$\mathcal{X}_i(t_j) = a_i \cos(2t_j) + b_i \sin(4t_j) + c_i(t_j^2 - \pi t_j + \frac{2}{9}\pi^2), \tag{6}$$

where $0 = t_1 < t_2... < t_{100} = \pi$ are equispaced points, a_i, b_i and c_i are independently drawn from a uniform distribution on $[0, 1]$, and n represents the sample size. Fig. 2 shows the simulated 215 curves for one replication.
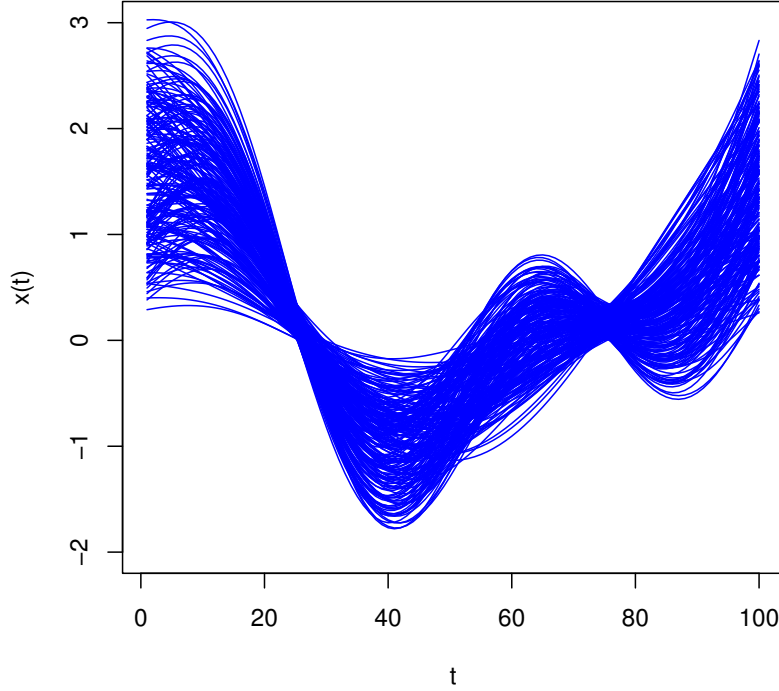

Figure 2: A sample of 215 simulated curves.

The multivariate response is simulated via the following steps.

1. Construct two regression function operators r_1 and r_2, each for one output. The two nonlinear operators r_1 and r_2 are defined as follows:

$$\begin{cases} r_1(\mathcal{X}_i) = \int_0^\pi t \cos(t)(\mathcal{X}_i'(t))^2 dt + \eta_i + \gamma_i, \\ r_2(\mathcal{X}_i) = \int_0^\pi [t \cos(t)(4b_i \cos(4t) - a_i \sin(t)) + c_i(2t - \pi)^2] dt + \eta_i + \gamma_i + \omega_i, \end{cases}$$

8

where $\mathcal{X}'(t)$ denotes the first derivative of $\mathcal{X}(t)$, η_i is a real-valued continuous variable following a standard normal distribution, γ_i is a discrete-valued variable drawn from a Bernoulli distribution, and ω_i is a real-valued continuous variable from an exponential distribution with rate parameter 1.

2. Generate two sets of random errors $\{\varepsilon_{1i}\}_{i=1,...,n}$ and $\{\varepsilon_{2i}\}_{i=1,...,n}$:

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} (\sigma_1)^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & (\sigma_2)^2 \end{bmatrix} \right),$$

where $\sigma_1 = 1$, $\sigma_2 = 2$, and we consider two cases for ρ: $\rho = 0.9$ (Case I) and $\rho = 0.1$ (Case II).

3. Compute the corresponding response variables:

$$Y_{1i} = r_1(\mathcal{X}_i) + \varepsilon_{1i}, \quad Y_{2i} = r_2(\mathcal{X}_i) + \varepsilon_{2i}, \quad i = 1, 2, ..., n.$$

We divide the 215 samples into two subsets: the first 160 are used for building the model, and the remaining 55 are used for prediction to assess the performance of the models. We compare the two proposed methods (MRC and MRD) with the independent response (IR) method where the two responses are modelled independently, based on the criterion of the root mean square prediction error defined as

$$\text{RMSPE} = \sqrt{MSE}, \quad MSE = \frac{1}{55} \sum_{i=161}^{215} (y_i - \hat{y}_i)^2.$$

The optimal bandwiths in MRC model and MRD model are determined by using a cross-validation procedure, that is, by minimising $CV(\mathbf{H}) = CV_1(h_1) + CV_2(h_2)$. While this minimisation is time-consuming, we see in our simulation studies that the optimal bandwidths for MRC, MRD and IR are quite close to each other, therefore, to simplify the computation we instead use a two-step CV procedure as done in Xiang et al. (2013): we first determine the individual bandwidths $(h_{1,0}, h_{2,0})$ independently, then determine the two bandwidths by minimising $CV(\mathbf{H})$ in a small neighbourhood of $(h_{1,0}, h_{2,0})$. In all the models considered, the Epanechnikov kernel function and semi-metric based on the second derivative are used. The above is repeated 10 times and the results are shown in Table 1. From Table 1, it can be seen that in case I, the MRC model performs better than the MRD model and IR model, while the MRD model performs slightly better than the IR model. On the other hand, neither MRC nor MRD are better than the IR method for Case II. This means that when the correlation is strong among different components of the response, the multivariate methods perform better than the independent model, even when the covariance is ignored in the model.

Table 1: Average RMSPE comparison between MRC, MRD and IR models based on ten replicated simulations.

Case	Components	RMSPE		
		MRC	MRD	IR
I	1	1.0041	1.0044	1.0210
	2	1.9678	1.9909	1.9937
II	1	1.0257	1.0381	1.0263
	2	2.1027	2.1440	2.0700

Example 2. In the second example we generate the functional predictor by

$$\mathcal{X}_i(t_j) = a_i(t_j - 0.5)^2 + b_i, \quad i = 1, 2, \dots, n = 100, \tag{7}$$

where $0 = t_1 < t_2 ... < t_{100} = 1$ are equispaced points, a_i and b_i are independently drawn from a uniform distribution on $[0,1]$, and n represents the sample size. Fig. 3 shows the simulated 100 curves from one replication.
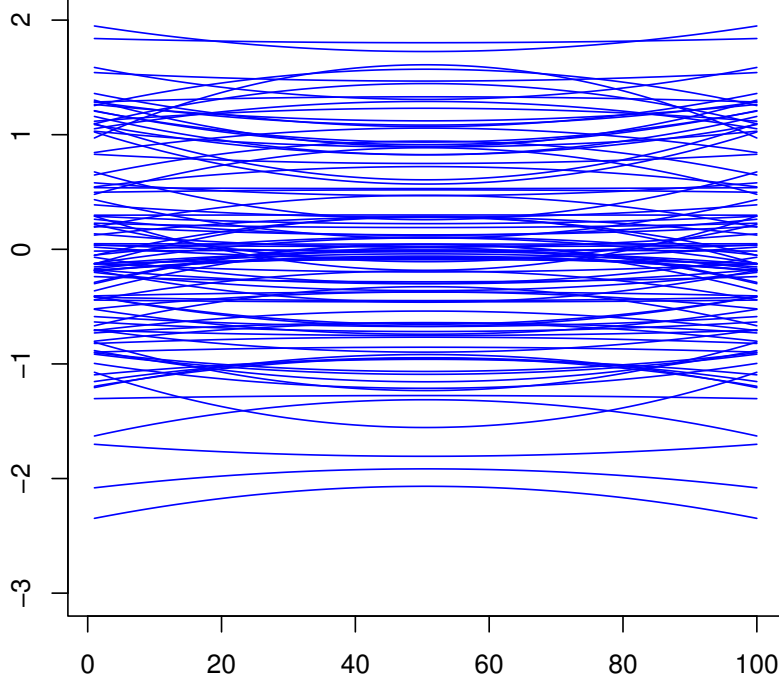

Figure 3: A sample of 100 simulated curves.

The regression functional operators r_1 and r_2 are constructed by

$$\begin{cases} r_1(\mathcal{X}_i) = \int_0^{\pi/2} \mathcal{X}_i'(t) dt, \\ r_2(\mathcal{X}_i) = \int_0^{\pi/2} |\mathcal{X}_i'(t)| \log|\mathcal{X}_i'(t)| dt. \end{cases}$$

The other settings including the random errors are the same as in Example 1.

We also divide the samples into two subsets: the training set contains the first 75 units and the testing set includes the remaining 25 units. We compare the two proposed models, the MRC and MRD models, with the IR model. The bandwiths are determined in the same way as before. The results based on 10 replications are reported in Table 2. These results confirm the conclusion of Example 1 that when the correlation is strong among different components of the response, the multivariate methods perform better than the independent model, whilst all the three methods give essentially the same results when the correlation is weak.

Table 2: Average RMSPE comparison between MRC, MRD and IR models based on ten replicated simulations.

		RMSPE		
Case	Components	MRC	MRD	IR
I	1	0.9900	0.9908	0.9936
	2	1.9750	1.9758	1.9858
II	1	1.0150	1.0154	1.0114
	2	1.9047	1.9049	1.9046

Example 3. We now consider another simulated example where the time points are randomly collected and the error term follows a multivariate t-distribution.

We generate the functional predictor by

$$\mathcal{X}_i(t_j) = a_i \cos(t) + b_i \sin(t) + W_i, \quad i = 1, 2, ..., 215, \tag{8}$$

where $0 = t_1 < t_2 ... < t_{100} = \pi$ are uniformly sampled in $[0, \pi]$, a_i, b_i and W_i are independently drawn from a uniform distribution on $[0.2, 1.5]$. Fig. 4 shows the simulated 215 curves from one replication.
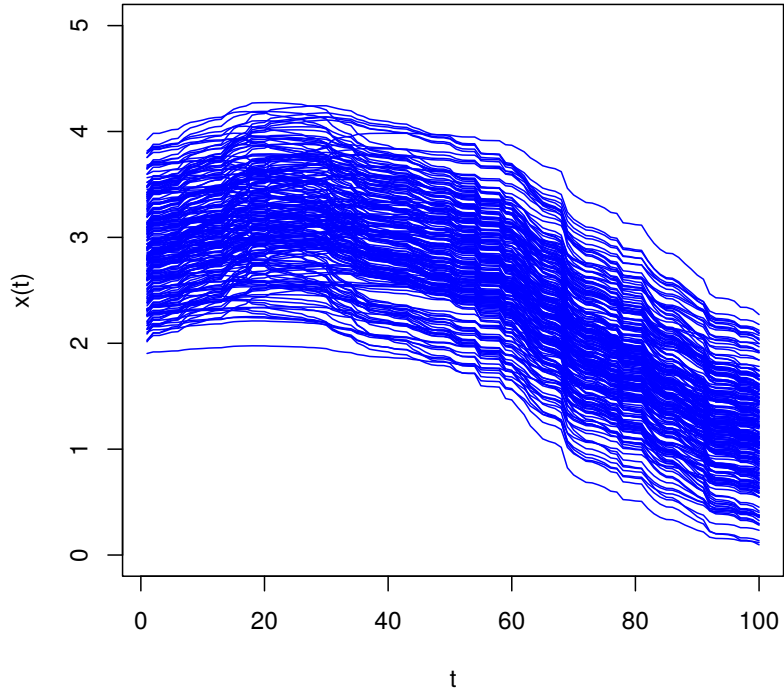

Figure 4: A sample of 215 simulated curves.

The two regression functional operators r_1 and r_2 are defined as: for $i = 1, ..., 215$

$$\begin{cases} r_1(\mathcal{X}_i) = \frac{1}{100} \int_0^\pi \mathcal{X}_i^2(t) dt, \\ r_2(\mathcal{X}_i) = \frac{1}{100} \int_0^\pi \mathcal{X}_i'(t) \mathcal{X}_i(t) dt, \end{cases}$$

where $\mathcal{X}'(t)$ denotes the first derivative of $\mathcal{X}(t)$. The response variables are given by

$$Y_{1i} = r_1(\mathcal{X}_i) + \varepsilon_{1i}, \quad Y_{2i} = r_2(\mathcal{X}_i) + \varepsilon_{2i}, \quad i = 1, 2, ..., 215,$$

11

where the error term $(\varepsilon_{1i}, \varepsilon_{2i})^t$ follows a bivariate t-distribution with the degrees of freedom $\nu = 3$ and the covariance matrix

$$\frac{\nu}{\nu - 2} \begin{bmatrix} (\sigma_1)^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & (\sigma_2)^2 \end{bmatrix},$$

where $\sigma_1 = 1$, $\sigma_2 = 2$, and we consider two cases for ρ: $\rho = 0.9$ (Case I) and $\rho = 0.2$ (Case II).

The same experiment procedure as in the first example is then conducted, and the average RMSPEs based on 10 repetitions by the three models are reported in Table 3. The results further confirm the conclusion of the previous two examples, that is, when the correlation is strong the MRC model significantly improves the prediction accuracy compared with the MRD and IR models, whilst all the three methods have similar performance when the correlation is weak.

Table 3: Average RMSPE comparison between the MRC, MRD and IR models based on ten replicated simulations.

Case	Components	RMSPE		
		MRC	MRD	IR
I	1	1.6195	1.6360	1.6373
	2	3.0334	3.0884	3.0663
II	1	1.4706	1.4689	1.4735
	2	3.0142	3.0004	3.0161

4.2   Real data example

We now demonstrate the effectiveness of the proposed methods using a real data example - the spectrometric data.

Spectrometric data is a well known example in functional data analysis, and several implementations have been undertaken using different models; see, for example, Ferraty and Vieu (2003, 2006), Ferraty et al. (2007), Benhenni et al. (2007), and Burba et al. (2009). The data comes from quality control, and can be found at *http://lib.stat.cmu.edu/datasets/tecator*. It concerns a sample of finely chopped pieces of meat. For each meat sample, the data consists of a 100-channel absorbance spectrum with a wavelength range between 850-1050nm (see Ferraty and Vieu (2006) for more details). The goal of spectrometric analysis is to determine the proportion of specific chemicals' content because the analysis by chemistry programming would take a lot more time and be more expensive. For this particular data, we want to predict fat, water and protein content simultaneously using the spectrometric curves as shown in Fig. 1.

In this analysis, the response variable Y_i are the percentages of fat, water and protein content in each piece of meat and the predictor is the spectrometric curves. We split the original sample into two subsets: the learning sample includes the first 160 units, and the testing sample consists of the last 55 units. The sample correlations between fat and water, between fat and protein and between water and protein are -0.9881, -0.8604 and 0.8145, respectively. The measure of performance for different methods is achieved by calculating the root mean square prediction error in the testing sample. The results are reported in Table 4.

From Table 4, it can be clearly seen that the two new multivariate methods (MRC and MRD) perform better than the independent model. Compared to the independent method, their RMSPE values are smaller across two components of the response variable (fat and water) whilst all the three methods give quite similar results for protein. This suggests that the multivariate methods are more accurate by modelling all the response components simultaneously.

Table 4: RMSPE comparison between MRC, MRD and IR models.

Components	RMSPE		
	MRC	MRD	IR
Fat	2.144	2.147	2.317
Water	1.957	1.970	2.068
Protein	1.603	1.606	1.602

5 Conclusion

This paper proposes two new methods for nonparametric functional regression when the covariate is functional and the response is multivariate. The first solution is to directly extend the nonparametric method for univariate response to multivariate response with different bandwidths for different responses, and in the second solution the correlation among different responses is taken into account and incorporated into the model. The rate of almost complete convergence is presented under certain conditions. The numerical examples presented in the paper suggest that our methods can perform well, especially when the correlation between different components of the response is strong.

It is noted that when we choose different bandwidths for different components of the response by a cross-validation procedure, our numerical results perform reasonably well. Xiang et al. (2013) mentioned that the bandwidths chosen by this approach usually have a large variability. Therefore, the research is needed to provide a framework for a simple procedure to choose the bandwidths, and our method may not be particularly suitable for high-dimensional multivariate functional data because of the complexity in choosing different bandwidths for different components. Furthermore, applying the proposed methodologies to other functional regression estimators, such as the k-nearest neighbour kernel estimator (Burba et al, 2009) and functional local linear kernel estimator (Benhenni et al, 2007), is also one of the candidates for future work.

Acknowledgement

References

Barrientos-Marin, J., Ferraty, F., and Vieu, P. (2010). Locally modelled regression and functional data. *Journal of Nonparametric Statistics*, 22(5), 617–632.

Benhenni, K., Ferraty, F., Rachdi, M., and Vieu, P. (2007). Local smoothing regression with functional data. *Computational Statistics*, 22(3), 353–369.

Boj, E., Delicado, P., and Fortiana, J. (2010). Distance-based local linear regression for functional predictors. *Computational Statistics and Data Analysis*, 54(2), 429–437.

Burba, F., Ferraty, F., and Vieu, P. (2009). k-Nearest Neighbour method in functional nonparametric regression. *Journal of Nonparametric Statistics*, 21(4), 453–469.

Chaouch, M., and Laïb, N. (2013). Nonparametric multivariate L_1-median regression estimation with functional covariates. *Electronic Journal of Statistics*, 7, 1553–1586.

Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2011). Kernel regression with functional response. *Electronic Journal of Statistics*, 5, 159–171.

Ferraty, F., Mas, A., and Vieu, P. (2007). Nonparametric regression on functional data: inference and practical aspects. *Australian and New Zealand Journal of Statistics*, 49(3), 267–286.

Ferraty, F., Van Keilegom, I., and Vieu, P. (2012). Regression when both response and predictor are functions. *Journal of Multivariate Analysis*, 109, 10–28.

Ferraty, F., and Vieu, P. (2003). Curves discrimination: a nonparametric functional approach. *Computational Statistics and Data Analysis*, 44(1), 161–173.

Ferraty, F., and Vieu, P. (2004). Nonparametric models for functional data, with application in regression, timeseries prediction and curve discrimination. *Nonparametric Statistics*, 16(1-2), 111–125.

Ferraty, F., and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice.* Springer Science and Business Media.

Lian, H. (2012). Convergence of nonparametric functional regression estimates with functional responses. *Electronic Journal of Statistics*, 6, 1373–1391.

Matsui, H., Araki, Y., and Konishi, S. (2008). Multivariate regression modeling for functional data. *Journal of Data Science*, 6(3), 313–331.

Ramsay, J. O., and Silverman, B. W. (1997). *Functional Data Analysis.* Springer, New York.

Ramsay, J. O., and Silverman, B. W. (2005). *Functional Data Analysis.* Second ed., Springer, New York.

Wang, B., and Chen, T. (2015). Gaussian process regression with multiple response variables. *Chemometrics and Intelligent Laboratory Systems*, 142, 159-165.

Xiang, D., Qiu, P., and Pu, X. (2013). Nonparametric regression analysis of multivariate longitudinal data. *Statistica Sinica*, 23(2), 769–789. www.jstor.org/stable/24310361.