Computational Statistics and Data Analysis 168 (2022) 107400

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

www.elsevier.com/locate/csda

Flexible quantile contour estimation for multivariate functional data: Beyond convexity

Gaurav Agarwal [a], Wei Tu [b], Ying Sun [a,*], Linglong Kong [c]

[a] Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia
[b] Department of Public Health Sciences and Canadian Cancer Trials Group, Queen's University, Kingston, ON, Canada
[c] Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB T6G 2G1, Canada

ARTICLE INFO

Article history:
Received 15 August 2020
Received in revised form 5 September 2021
Accepted 12 November 2021
Available online 16 November 2021

Keywords:
Consistency
Functional data analysis
Multivariate quantile
Nonconvex
Quantile contour
$PM_{2.5}$

ABSTRACT

Nowadays, multivariate functional data are frequently observed in many scientific fields, and the estimation of quantiles of these data is essential in data analysis. Unlike in the univariate setting, quantiles are more challenging to estimate for multivariate data, let alone multivariate functional data. This article proposes a new method to estimate the quantiles for multivariate functional data with application to air pollution data. The proposed multivariate functional quantile model is a nonparametric, time-varying coefficient model, and basis functions are used for the estimation and prediction. The estimated quantile contours can account for non-Gaussian and even nonconvex features of the multivariate distributions marginally, and the estimated multivariate quantile function is a continuous function of time for a fixed quantile level. Computationally, the proposed method is shown to be efficient for both bivariate and trivariate functional data. The monotonicity, uniqueness, and consistency of the estimated multivariate quantile function have been established. The proposed method was demonstrated on bivariate and trivariate functional data in the simulation studies, and was applied to study the joint distribution of $PM_{2.5}$ and geopotential height over time in the Northeastern United States; the estimated contours highlight the nonconvex features of the joint distribution, and the functional quantile curves capture the dynamic change across time.

1. Introduction

Functional data, where each point is a curve, surface, or anything else varying over a continuum, arise in all areas of science, medicine, and engineering. During the last two decades, the area of functional data analysis (FDA) has greatly attracted attention among researchers. However, most of the methods available in the literature are developed for univariate functional data, while the analysis of multivariate functional data has been largely overlooked. Multivariate functional data can be viewed as realizations of multivariate random functions; for example, children's height and weight over age, patient's CD4 cell count and viral load over the duration of therapy, and observed variables such as temperature, geopotential height, and $PM_{2.5}$ over time. Computing quantiles is one of the essential tasks for data analysis; however, the extension of quantiles to multivariate functional data is quite challenging. In this paper, we propose a method to estimate the quantile contours of multivariate functional data.

* Corresponding author.
E-mail address: ying.sun@kaust.edu.sa (Y. Sun).

https://doi.org/10.1016/j.csda.2021.107400
0167-9473/

$PM_{2.5}$ is an air pollutant that represents the concentration of fine particulate matter with a diameter less than 2.5 μm. The exposure to high concentrations of $PM_{2.5}$ poses a considerable risk for human health and have known to cause respiratory and cardiovascular illness (World Health Organization, 2003; Cohen et al., 2005; Fuentes et al., 2006; Hu et al., 2008). Various meteorological variables, such as temperature, wind, geopotential height, affect the $PM_{2.5}$ concentrations significantly (Li et al., 2015; Russell et al., 2017). However, most of these analyses were done in a univariate regression framework, and the effects were estimated at the mean level without reflecting any dynamic change in time. Consequently, a deeper understanding of the relationship between $PM_{2.5}$ and meteorological variables can help us in the formulation of pollution control strategies and health-care policies. In this article, we study the joint distribution of $PM_{2.5}$ and geopotential height at 850 hPa over time across the Northeastern United States. The bivariate distribution of $PM_{2.5}$ and geopotential height at a given time may have nonconvex features and is of particular interest as high-pressure systems can trap pollutants and cause high $PM_{2.5}$ concentrations. Previous literature has shown that several air pollution crises have occurred when maximum values of geopotential height were recorded above the city (Saunders and Waugh, 2015; He et al., 2018). The proposed method is useful in applications where the marginal distribution is of interest which is observed over time. For the $PM_{2.5}$ and geopotential height data application, the scientific goal is to study the time-evolving nonconvex features of the bivariate distribution.

For univariate data, the natural linear ordering makes it convenient and straightforward to define quantiles. However, the generalization to multivariate data is not trivial at all, let alone functional data. Most of the methods in the literature rely on ranking multivariate or functional data using depth functions (Hettmansperger et al., 1992; Chaudhuri, 1996; Breckling et al., 2001; Chakraborty, 2001, 2003; Serfling, 2010; Chakraborty and Chaudhuri, 2014). For example, Kong and Mizera (2012) use projection-based quantiles to estimate multivariate quantile contours, which coincide with Tukey depth contours (Tukey, 1975); see also Hallin et al. (2010); Paindaveine and Šiman (2012); Liu et al. (2013). For univariate functional data, Serfling and Wijesuriya (2017) use spatial depth to estimate quantile curves, see also Chowdhury and Chaudhuri (2017, 2019). Sun and Genton (2011) defined functional boxplots based on functional band depth (López-Pintado and Romo, 2009). Fraiman and Pateiro-López (2012) defined projection-based quantiles for high dimensional multivariate data and univariate functional data. For multivariate functional data, Pietrosanu et al. (2021) modeled bivariate functional data using directional quantiles, and the estimated quantile contours for the marginal distribution coincides with the Tukey depth contours, see also (Agarwal and Sun, 2020). However, most existing depth-based contours are convex by nature, which is not desirable for distributions with nonconvex density contours, which we refer to as nonconvex distributions. Most of these depth-based methods follow linear monotonicity axiom, which imposes star-convexity of its contours (Liu, 1990; Zuo and Serfling, 2000). They also suffer from computational issues in higher dimensions, and even computing quantiles for three-dimensional data is very time-consuming and challenging (Zuo, 2019).

Wei (2008) proposed an approach based on stratified quantile regression to estimate bivariate quantile contours for longitudinal data. Their quantile contours can account for nonconvex distributions; however, their method relies on sequential conditioning of response components, and the order adopted for that conditioning. Recently, Chernozhukov et al. (2017) introduced a concept of multivariate quantiles based on deterministic maps that transform a reference distribution to the distribution of interest. Their approach gives up the linear monotonicity axiom and can pick up nonconvex features of the distribution. Carlier et al. (2016) also proposed a conditional multivariate quantile function and a multivariate quantile regression model using an optimal transport approach. However, multivariate data methods cannot be used for functional data because of the dimensionality issues. There has been a limited focus in the current literature on the estimation of quantiles for multivariate functional data. Multivariate functional data are actively generated in various research areas, including meteorology, biology, biomedicine, engineering, and environmental science. The estimated quantile contours for multivariate functional data can be used to better understand the joint distribution of the data and for tasks such as clustering and outlier detection.

To model the multivariate functional data, we consider a cubic spline model in time and represent it as a linear combination of B-spline basis functions. We consider a multivariate functional quantile regression framework to model the multivariate quantile function conditioned on the covariates. In this paper, we only consider models with intercept to define quantiles for multivariate functional data, but the proposed methodology can incorporate covariate information easily. The quantile contours are constructed from the estimated multivariate quantile function for a particular time point, and functional quantile curves are constructed over time for each $\boldsymbol{u}$, where $\boldsymbol{u}$ is the multivariate quantile index. The estimated multivariate quantile function follows the property of monotonicity and uniqueness, and consistency has been established. The estimated quantile contour can account for non-Gaussian and even nonconvex distributions, which the depth-based methods fail to do. The estimated multivariate quantile function is a continuous function of time for a fixed quantile index, which is useful to capture the change in the distribution over time. Computationally, the proposed method is also efficient for higher dimensions, while existing methods only deal with bivariate functional data.

In the simulation studies, we apply the proposed methods to bivariate and trivariate functional data. We show that in the case of Gaussian marginals distribution, the predicted quantile contours coincide with the density contours, and for nonconvex marginal distribution, the contours can correctly pick up the shape of the distribution, while the depth-based methods failed to do so. The prediction was identified to be reasonably accurate in predicting the center of the distribution. For the application, the bivariate distribution of $PM_{2.5}$ and geopotential height is estimated using flexible quantile contours for the northeastern United States across six months. The quantile contours explain the nonconvexity in the marginal distribution, and the functional quantile curves capture the dynamic change in the distribution over time.

The rest of the paper is organized as follows. In Section 2, we define the multivariate functional quantile model, discuss the estimation of multivariate functional quantiles and construction of quantile contours. We discuss the theoretical properties of the multivariate quantile function in Section 3. In Section 4, we conduct simulation studies on bivariate and trivariate functional data. In Section 5, we apply our methods to air pollution data, and, finally, Section 6 concludes the paper with a discussion.

2. Methodology

2.1. Multivariate functional quantile model

Let $Y(t)$ be a stochastic process defined over a continuum $t \in \mathcal{T}$ in $\mathbb{R}$. Let us assume that $Y \in L^2(\mathcal{T})$, where $L^2(\mathcal{T})$ is a Hilbert space of square-integrable functions on the interval $\mathcal{T}$. Multivariate functional data can be treated as realizations of multivariate random processes on a given interval. We consider multivariate functional data, $\boldsymbol{Y}(t) = (Y_1(t), \ldots, Y_d(t))^\top \in \mathbb{R}^d$, where $t \in \mathcal{T}$, and $\boldsymbol{X} \in \mathbb{R}^p$ as the covariates. Let $F_{\boldsymbol{X}}$ denote the distribution of covariates $\boldsymbol{X}$ and $\mathcal{X}$ the support of $F_{\boldsymbol{X}}$. Let $\boldsymbol{U}$ be a latent vector that follows a fixed reference distribution $F_{\boldsymbol{U}}$, for instance uniform distribution on $(0,1)^d$. Let $\mathcal{U}$ and $\mathcal{Y}$ denote the support of $F_{\boldsymbol{U}}$, and $F_{\boldsymbol{Y}(t)}$, respectively. The idea is to create a deterministic map from the random vector $\boldsymbol{U}$, with reference distribution $F_{\boldsymbol{U}}$ to the vector of interest $\boldsymbol{Y}(t)$ conditional on $\boldsymbol{X}$. We assume that the reference distribution $F_{\boldsymbol{U}}$ has a probability density function $f_{\boldsymbol{U}}$ with respect to the Lebesgue measure on $\mathbb{R}^d$ and has convex support on $\mathcal{U}$. Let $(\boldsymbol{Y}(t), \boldsymbol{X})$ and $\boldsymbol{U}$ be defined on the complete probability space $(\Omega, \mathcal{A}, P)$. Denote by $F_{\boldsymbol{Y}(t)\boldsymbol{X}}$ the joint distribution of $(\boldsymbol{Y}(t), \boldsymbol{X})$, and $F_{\boldsymbol{Y}(t)|\boldsymbol{X}}$ the conditional distribution of $\boldsymbol{Y}(t)$ given $\boldsymbol{X}$. The multivariate functional quantile model is the following linear regression model given by

$$\boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{U}, \boldsymbol{X}) = \boldsymbol{\beta}(\boldsymbol{U}, t)^\top \boldsymbol{X}, \quad \boldsymbol{U} \sim F_{\boldsymbol{U}}, \quad \mathrm{E}(\boldsymbol{X} \mid \boldsymbol{U}) = \mathrm{E}(\boldsymbol{X}), \tag{1}$$

where $\boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u}, \boldsymbol{x})$ is the conditional multivariate quantile function, and $\boldsymbol{u} \mapsto \boldsymbol{\beta}(\boldsymbol{u}, t)^\top \boldsymbol{x}$ is a monotone map from $\mathcal{U}$ to $\mathbb{R}^d$, in the sense of being a gradient of a convex function:

$$\boldsymbol{\beta}(\boldsymbol{u}, t)^\top \boldsymbol{x} = \nabla_{\boldsymbol{u}} \boldsymbol{B}(\boldsymbol{u}, t)^\top \boldsymbol{x}, \quad \text{for all } \boldsymbol{u} \in \mathcal{U},\ \boldsymbol{x} \in \mathcal{X}, \text{ and } t \in \mathcal{T}, \tag{2}$$

where $\boldsymbol{u} \mapsto \boldsymbol{B}(\boldsymbol{u}, t)^\top \boldsymbol{x}$ is a strictly convex map from $\mathcal{U}$ to $\mathbb{R}$. Since we want the multivariate quantile function to have a monotonicity property (defined in Section 3), we require it to be a gradient of a convex function. The existence of a convex function with the required properties is established by the fundamental results in McCann (1995). The parameter $\boldsymbol{\beta}(\boldsymbol{u}, t)$ of the model is a continuous function of t, and is indexed by the multivariate quantile index $\boldsymbol{u} \in \mathcal{U}$. For a time t, $\boldsymbol{\beta}(\boldsymbol{u}, t)$ is a $d \times p$ matrix of regression coefficients, which reduces to a vector of quantile regression coefficients for $d = 1$. The proposed multivariate functional quantile model is a nonparametric, time-varying coefficient model where $\boldsymbol{B}(\boldsymbol{u}, t)$ are smooth functions of t. The time-varying coefficients can be well approximated by basis function expansions (Huang et al., 2002). We choose B-splines expansion as they have compact support and are computationally efficient. The number of knots are fixed and the knots are typically chosen as suitable quantiles of the observed set of times t. Let $\boldsymbol{H}(t) = (h_1(t), \ldots, h_K(t))^\top$ be a B-spline basis, where K is the number of B-spline functions, then $\boldsymbol{B}(\boldsymbol{u}, t)$ is approximated as $\boldsymbol{B}(\boldsymbol{u}, t) = \boldsymbol{\gamma}(\boldsymbol{u})\boldsymbol{H}(t)$, where $\boldsymbol{\gamma}(\boldsymbol{u})$ is a $p \times K$ matrix of coefficients. Then the equation (2) can be rewritten as

$$\boldsymbol{\beta}(\boldsymbol{u}, t)^\top \boldsymbol{x} = \nabla_{\boldsymbol{u}} \{\boldsymbol{\gamma}(\boldsymbol{u})\boldsymbol{H}(t)\}^\top \boldsymbol{x}, \quad \text{for all } \boldsymbol{u} \in \mathcal{U},\ \boldsymbol{x} \in \mathcal{X}, \text{ and } t \in \mathcal{T}. \tag{3}$$

The conditional multivariate quantile function, $\boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u}, \boldsymbol{x})$, is a monotone map $\boldsymbol{u} \mapsto \boldsymbol{\beta}(\boldsymbol{u}, t)^\top \boldsymbol{x}$, that transforms the vector $\boldsymbol{U}$ into $\boldsymbol{Y}(t)$ conditional on $\boldsymbol{X}$. It is a continuous function of t, for a fixed $\boldsymbol{u}$. In this paper, we consider only the intercept term in the model to obtain the multivariate quantiles of $\boldsymbol{Y}(t)$. In principle, we can consider more covariates to study the multivariate quantiles of $\boldsymbol{Y}(t)$ conditioned on the covariates. In the univariate case, $\boldsymbol{U}$ is interpreted as a notion of rank, and the typical choice of the reference distribution is Uniform$(0,1)$. In the multivariate case, $\boldsymbol{U}$ can be interpreted as a multivariate rank of $\boldsymbol{Y}(t)$, with a typical choice of Uniform$(0,1)^d$, since we create a deterministic map from $\boldsymbol{U}$ to $\boldsymbol{Y}(t)$. The d-dimensional vector $\boldsymbol{U}$ we used is the same as in Carlier et al. (2016), so the associated quantiles are d-dimensional marginal quantiles. The quantile functions are modeled by spline regression models. For d-variate functions, $\boldsymbol{U}$ is the marginal quantile index in $\mathbb{R}^d$, which can be viewed as the quantile direction for d-dimensional reference distribution marginally, or for a given t. Other choices of reference distributions include the uniform distribution in the unit ball, the standard normal distribution $N(0,1)^d$. We choose Uniform$(0,1)^d$ as the reference distribution so that the proposed model becomes analogous to the classical univariate quantile regression for scalar $Y(t)$ at time t. The conditional multivariate quantile function can be interpreted as the optimal transport from the reference distribution $F_{\boldsymbol{U}}$ to $F_{\boldsymbol{Y}(t)|\boldsymbol{X}}$. There might be other maps $\boldsymbol{Q} : (\mathcal{U}, \mathcal{X}) \mapsto \mathcal{Y}$, such that if $\boldsymbol{V} \sim F_{\boldsymbol{U}}$, then $\boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{V}, \boldsymbol{X})$ defines a transport from $F_{\boldsymbol{U}}$ to $F_{\boldsymbol{Y}(t)|\boldsymbol{X}}$. But the objective is to take the optimal transport such that it has the minimum Wasserstein distance $\mathrm{E}\|\boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{V}, \boldsymbol{X}) - \boldsymbol{V}\|^2$, where $\|.\|$ denotes the Euclidian norm of $\mathbb{R}^d$. Note that unlike the scalar case, we can not compute the quantile $\boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u}, \boldsymbol{x})$ for any given $\boldsymbol{u}$ without first computing the whole quantile map $\boldsymbol{u} \mapsto \boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u}, \boldsymbol{x})$.

2.2. Estimation of conditional multivariate functional quantiles

To estimate the conditional multivariate quantile function, we estimate the regression coefficients of the multivariate functional quantile model (1) following the approach of Carlier et al. (2016) based on optimal transport. We assume that the second moment of $\boldsymbol{Y}(t)$ and second moment of $\boldsymbol{V}$ are finite. We minimize the Wasserstein distance between $\boldsymbol{V}$ and $\boldsymbol{Y}(t)$ considered over all random vectors $\boldsymbol{V}$ defined in the probability space $(\Omega, \mathcal{F}, P)$ subject to mean independence constraint,

$$\min_{\boldsymbol{V}}\{\mathrm{E}(\|\boldsymbol{V}-\boldsymbol{Y}(t)\|^2) : \boldsymbol{V} \sim F_{\boldsymbol{U}}, \quad \mathrm{E}(\boldsymbol{X} \mid \boldsymbol{V}) = \mathrm{E}(\boldsymbol{X})\}.$$

This problem is the conditional version of the Monge-Kantorovich optimal transport problem with Brenier's quadratic costs which was solved by Brenier (1991). The problem is equivalent to the following covariance maximization problem

$$\max_{\boldsymbol{V}}\{\mathrm{E}(\boldsymbol{V}^\top \boldsymbol{Y}(t)) : \boldsymbol{V} \sim F_{\boldsymbol{U}}, \quad \mathrm{E}(\boldsymbol{X} \mid \boldsymbol{V}) = \mathrm{E}(\boldsymbol{X})\}. \tag{4}$$

For the estimation of conditional multivariate functional quantiles, we simply need to solve the optimization problem (4), which conveniently leads to a dual problem and leads to convex analysis. The solution to the optimization problem (4) exist and is given by $\boldsymbol{V} = \boldsymbol{U}$, where $\boldsymbol{U} = \boldsymbol{Q}^{-1}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{Y}(t), \boldsymbol{X})$, which follows from Theorem 2.3 of Carlier et al. (2016). So, when we fully specify the reference distribution $F_{\boldsymbol{U}} = \mathrm{Uniform}(0,1)^d$, then the vector $\boldsymbol{U}$ measures the centrality of the observation $\boldsymbol{Y}(t)$ for each of the dimensions, conditional on $\boldsymbol{X}$. The vector $\boldsymbol{U}$ is generated from fully specified reference distribution $F_{\boldsymbol{U}}$, which serves as the d-dimensional quantile indices. The realizations are treated as fixed. The following paragraph gives the technical details of the implementation in practice.

Multivariate functional data is typically recorded at discrete points. Consider discretized multivariate functional data $\boldsymbol{Y}(t)$, $t = t_1, \ldots, t_r$, with m observations at each time point. Let $\mathrm{Y} = \{\mathrm{Y}(t_1)^\top, \ldots, \mathrm{Y}(t_r)^\top\}^\top$ be a $mr \times d$ matrix, where $\mathrm{Y}(t_i)$ is $m \times d$ matrix with each row as a realization of $\boldsymbol{Y}(t_i)$, $i = 1, \ldots, r$. Here, we consider only the intercept in our model, so X is the $mr \times K$ B-spline matrix. We generate n points for $\boldsymbol{U}$ from the fixed reference distribution Uniform $(0,1)^d$. Let U be a $n \times d$ matrix with each row as a realization of $\boldsymbol{U}$. The distributions F_{YX} and F_{U} can be approximated by discrete distributions ν and μ. Let ν_j be the probability attached with (y_j, x_j), $j = 1, \ldots, mr$ and $\nu = (\nu_1, \ldots, \nu_{mr})^\top$. Let μ_i be the probability attached with u_i, $i = 1, \ldots, n$, and $\mu = (\mu_1, \ldots, \mu_n)^\top$. Let p_{ij} denote the probability attached with (u_i, x_j, y_j), then the objective is to find the $n \times mr$ optimal matrix $\mathrm{p} = ((p_{ij}))$ such that it maximizes the objective function

$$\sum_{ij} p_{ij} y_j^\top u_i = \mathrm{tr}(\mathrm{U}^\top \mathrm{pY}),$$

with respect to constraints $\mathrm{p}^\top 1_n = \nu$ and $\mathrm{pX} = \mu\nu^\top \mathrm{X}$. This constrained optimization problem is solved using generalized Lagrange function and the primal problem is given by

$$\max_{\mathrm{vec(p)}\geq 0} \mathrm{vec}\Big(\mathrm{UY}^\top\Big)^\top \mathrm{vec(p)} : \quad \Big(I_{mr} \otimes 1_n^\top\Big)\mathrm{vec(p)} = \nu, \Big(\mathrm{X}^\top \otimes I_n\Big)\mathrm{vec(p)} = \mathrm{vec}\Big(\mu\nu^\top \mathrm{X}\Big). \tag{5}$$

Using generalized Lagrange function, the dual form can be derived as

$$\min_{\psi, \mathrm{b}} \psi^\top \nu + \mathrm{vec(b)}^\top \mathrm{vec}(\mu\nu^\top \mathrm{X}) : \quad \Big(I_{mr} \otimes 1_n\Big)\psi + \Big(\mathrm{X} \otimes I_n\Big)\mathrm{vec(b)} \geq \mathrm{vec}(\mathrm{UY}^\top), \tag{6}$$

where ψ is $mr \times 1$ vector and b is a $n \times K$ matrix. The proof of obtaining the dual form (6) from the primal form (5) is given in the Appendix C. The matrices $\Big(I_{mr} \otimes 1_n\Big)\psi$ and $\Big(\mathrm{X} \otimes I_n\Big)\mathrm{vec(b)}$ can be very large matrices but they are sparse. This sparsity can be used for computational advantages. The above problems can be solved by linear programming software, such as Gurobi in R.

2.3. Quantile contour estimation and prediction

The multivariate quantile function can adequately capture the features of a wide range of distributions. In this section, we describe the estimation of quantile contours at time t, and functional quantile curves over time from the predicted multivariate quantile function. The optimal solution of matrix $\mathrm{b}_{n\times K}$ from the dual problem (6) gives us the estimated regression coefficients $\boldsymbol{b}(\boldsymbol{u}, t) = \boldsymbol{B}(\boldsymbol{u}, t)$ and, hence, gives the convex map $\boldsymbol{u} \mapsto \boldsymbol{B}(\boldsymbol{u}, t)^\top \boldsymbol{x}$. The map $\boldsymbol{u} \mapsto \boldsymbol{\beta}(\boldsymbol{u}, t)^\top \boldsymbol{x}$ is obtained by taking the gradient of convex function $\boldsymbol{B}(\boldsymbol{u}, t)^\top \boldsymbol{x}$. The multivariate quantile function in time, $\hat{\boldsymbol{Q}}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u}, \boldsymbol{x})$, is then obtained as a convex map from the reference distribution. The components of the multivariate quantile function are given by

$$\hat{\boldsymbol{Q}}_{Y_1(t)|\boldsymbol{X}}(u_1,\ldots,u_d) = \boldsymbol{\beta}_1(u_1,\ldots,u_d,t)^\top \boldsymbol{x} = \frac{\partial \boldsymbol{B}}{\partial u_1}(u_1,\ldots,u_d,t)^\top \boldsymbol{x}$$

$$\vdots$$

$$\hat{\boldsymbol{Q}}_{Y_d(t)|\boldsymbol{X}}(u_1,\ldots,u_d) = \boldsymbol{\beta}_d(u_1,\ldots,u_d,t)^\top \boldsymbol{x} = \frac{\partial \boldsymbol{B}}{\partial u_d}(u_1,\ldots,u_d,t)^\top \boldsymbol{x}.$$

After obtaining the multivariate quantiles at time t, we estimate the marginal quantile contours. To construct the quantile contours, we first map the reference distribution Uniform $(0,1)^d$ to a spherical uniform distribution with an optimal assignment algorithm. For $d = 2$, the spherical uniform distribution is generated randomly on circles with evenly spaced radii in (0,1]. An assignment problem is a matching problem that minimizes the total cost of the assignment; in our case, the cost is the Euclidean distance. Since the number of points in the reference distribution and spherical uniform distribution are equal, it is a linear assignment problem. The assignment problem can be solved using algorithms such as the linear sum assignment algorithm (Burkard et al., 2009), Hungarian algorithm (Papadimitriou and Steiglitz, 1982), or an adaptation of simplex algorithm (Murty, 1983).

Once we obtain the map between the reference distribution and the spherical uniform distribution, we compute the norm of each point in the spherical uniform distribution. Finally, we construct α-hulls of $\hat{\boldsymbol{Q}}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\mathcal{U}_n \cap \mathbb{S}(\tau))$, where $\mathcal{U}_n$ is the set of n points generated from the reference distribution and $\mathbb{S}(\tau) = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| \leq \tau\}$ for $\tau \in (0,1]$. The α-hull is the τ-quantile region with probability mass τ, and the boundary of the α-hull is the τ-quantile contour. The algorithm of α-hull is based on Delaunay triangulation and is explained in Edelsbrunner et al. (1983), and the boundary of α-hulls are formed by the arc of open balls of radius α. The proposed quantile contours are not restricted with convexity and work well for non-Gaussian, and even nonconvex distributions, without making any distributional assumptions. The functional quantile curves are also estimated using the multivariate quantile function to study the pattern over time. For a fixed $\boldsymbol{u}$, the multivariate quantile function, $\hat{\boldsymbol{Q}}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u},\boldsymbol{x})$, is a continuous function of t. The functional quantile curves are constructed by $\hat{\boldsymbol{Q}}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u},\boldsymbol{x})$ over t, for each fixed $\boldsymbol{u}$. The median functional quantile curve is also estimated to summarize the pattern over time, which is $\hat{\boldsymbol{Q}}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u}_{\mathrm{m}},\boldsymbol{x})$ over t, where $\boldsymbol{u}_{\mathrm{m}}$ is the component-wise median of the reference distribution. However, if we had neglected the functional nature of data and modeled the multivariate quantile function without the spline representation, then the estimated quantile contours would be rougher for each time point, and it would not be able to account for changes over time. Here, the proposed method borrows strength from other time points, and the functional quantile curves are well defined over time.

3. Theoretical properties

The conditional multivariate functional quantile has some desired properties that are analogous to the classical conditional quantile function:

Property 1 *(Monotonicity).* The conditional multivariate functional quantile is a gradient of a convex function. The gradient of a convex function is monotone (Minty, 1964). Therefore, the conditional multivariate quantile function follows monotonicity property with respect to $\boldsymbol{u} \in \mathcal{U}$:

$$\{\boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u},\boldsymbol{x}) - \boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\bar{\boldsymbol{u}},\boldsymbol{x})\}^\top (\boldsymbol{u} - \bar{\boldsymbol{u}}) \geq 0, \quad \text{for all } \boldsymbol{u}, \bar{\boldsymbol{u}} \in \mathcal{U},\ t \in \mathcal{T}.$$

Monotonicity is an evident feature of univariate quantile functions, and it is more complicated to define monotonicity properties in the multivariate case. We require the multivariate quantile function $\boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u},\boldsymbol{x})$ to be the gradient of a convex function, which is a plausible generalization of monotonicity for the case of scalar $Y(t)$. In the scalar case, the requirement that the function is the gradient of the convex function reduces to that function is non-decreasing. It implies that if we increase a particular variable in $\boldsymbol{U}$ while keeping others fixed, then the corresponding variable in $\boldsymbol{Y}(t)$ is non-decreasing.

Property 2 *(Uniqueness).* The conditional multivariate quantile function $\boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u},\boldsymbol{x})$ is a unique gradient of convex function, $F_{\boldsymbol{U}}$-almost everywhere. The gradient $\nabla_{\boldsymbol{u}} \boldsymbol{B}(\boldsymbol{u},t)^\top \boldsymbol{x}$ exists $F_{\boldsymbol{U}}$-almost everywhere, i.e., the set of points where it does not exist is negligible.

This property implies that the map $\boldsymbol{u} \mapsto \boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u},\boldsymbol{x})$ exists and is the unique gradient of convex function such that, whenever $\boldsymbol{V} \sim F_{\boldsymbol{U}}$, the random vector $\boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{V},\boldsymbol{x})$ has the distribution of $\boldsymbol{Y}(t)$ conditional on $\boldsymbol{X} = \boldsymbol{x}$. This map is unique in the sense that if there is any other map with the same properties of $\boldsymbol{Q}$, then it will agree with it $F_{\boldsymbol{U}}$-almost everywhere (see Villani, 2003).

The above properties immediately follow from the properties of the multivariate conditional quantile function defined by Carlier et al. (2016).

Theorem 1 *(Uniform convergence). Let $\mathcal{U}$ denote the compact subset of $\mathbb{R}^d$ and the support of the reference distribution $F_{\boldsymbol{U}}$. Let $\hat{\boldsymbol{Q}}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u})$ be the predicted conditional multivariate quantile function on $\mathcal{U}$. We denote the number of points generated from the reference distribution as n. Under the regularity conditions that the sample functions belong to a Hilbert space of square-integrable functions on the interval $\mathcal{T}$ in $\mathbb{R}$ and the number of sample functions are finite, as $n \to \infty$*

$$\sup_{\boldsymbol{u}\in\mathcal{U}} \| \hat{\boldsymbol{Q}}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u}) - \boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u}) \| \to 0.$$

$\hat{\boldsymbol{Q}}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u})$ uniformly converges to $\boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u})$ $F_{\boldsymbol{U}}$-almost everywhere, i.e., the set where it does not converge is negligible.

Corollary 1. *Suppose A is any non-empty compact subset of $\mathcal{U}$, and denoting $\hat{\boldsymbol{Q}}(A) = \{\hat{\boldsymbol{Q}}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u}); \boldsymbol{u} \in A\}$ and $\boldsymbol{Q}(A) = \{\boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u}); \boldsymbol{u} \in A\}$, then as $n \to \infty$*

$$\sup_{A\subseteq\mathcal{U}} d_H(\hat{\boldsymbol{Q}}(A), \boldsymbol{Q}(A)) \to 0,$$

where d_H is the Hausdorff distance.

The proofs of Theorem 1 and Corollary 1 are given in Appendix A and Appendix B, respectively. Theorem 1 establishes the uniform consistency of the predicted multivariate quantile function $\hat{\boldsymbol{Q}}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u})$ to $\boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u})$ $F_{\boldsymbol{U}}$-almost everywhere. It implies that as the sample size of the reference distribution increases, the predicted conditional quantile function uniformly converges to the theoretical multivariate quantile function. Note that the actual sample of functions are observed only at discrete time points. The quantile function is estimated from the set of observed values of sample functions at discrete time points using the spline representation in the model (1) and is a continuous function of $t \in \mathcal{T}$. The set $\boldsymbol{Q}(A)$ with $A = (\mathcal{U}_n \cap \mathbb{S}(\tau))$ and $A = (\mathcal{U}_n \cap \mathcal{S}(\tau))$ is the τ-quantile region and τ-quantile contours, respectively, where $\mathcal{S}(\tau) = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| = \tau\}$. Therefore, Corollary 1 establishes the consistency of the predicted quantile region and contours to the theoretical quantile region and contours..

4. Simulation study

4.1. Bivariate functional data with convex margins

To illustrate our methods, we first conduct simulation studies by generating bivariate functional data with convex marginal distributions. Here, we simulate bivariate functional data $\boldsymbol{Y}(t) = (Y_1(t), Y_2(t))^\top$ with bivariate normal error distribution using the following model

$$\begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1^\top \boldsymbol{\theta}(t) \\ \boldsymbol{X}_2^\top \boldsymbol{\theta}(t) \end{pmatrix} + \begin{pmatrix} \epsilon_1(t) \\ \epsilon_2(t) \end{pmatrix}, \quad \text{where} \begin{pmatrix} \epsilon_1(t) \\ \epsilon_2(t) \end{pmatrix} \sim N\left[\begin{pmatrix} 5 \\ 5 \end{pmatrix}, \quad \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix} \right],$$

and $\boldsymbol{X}_1^\top \boldsymbol{\theta}(t)$ and $\boldsymbol{X}_2^\top \boldsymbol{\theta}(t)$ represents cubic spline functions in t. We represent the spline functions as a linear combination of B-spline basis functions with the number of knots equal to 3. We set the two coefficients as $\boldsymbol{X}_1 = (1, 0.7, 0.8, 0.9, 0.7, 0.8, 0.9)^\top$ and $\boldsymbol{X}_2 = (1, 0.8, 0.7, 0.8, 0.9, 0.7, 0.6)^\top$, and $\boldsymbol{\theta}(t) = (h_1(t), \ldots, h_7(t))^\top$ is a cubic B-spline basis including intercept with three knots. The data was generated for $m = 300$ and 10 equispaced points in [0,1].

We then use the model (1) to estimate the conditional multivariate quantile function with reference distribution $U(0,1)^2$. We generate the points of the reference distributions from a 21×21 uniformly spaced grid of points with $n = 441$. We then solve the linear programming problem (6) using Gurobi in R, and estimate the conditional multivariate quantile function. The estimation of the model was not sensitive to the number of knots as the value of the minimum objective function in (6) was stable with an increase in the number of knots.

The quantile contours are estimated for the marginal distributions of the simulated data, and Fig. 1a shows the estimated quantile contours for $t = 0.48$, for three quantile levels, 0.25, 0.5, and 0.75. In the data generating model, the errors are simulated from a bivariate normal distribution, hence the marginal distribution is known and given by

$$\begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{X}_1^\top \boldsymbol{\theta}(t) + 5 \\ \boldsymbol{X}_2^\top \boldsymbol{\theta}(t) + 5 \end{pmatrix}, \quad \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix} \right].$$

The density contours of the bivariate normal distribution are also plotted for $t = 0.48$ (Fig. 1a) at the levels 0.25, 0.5, and 0.75. It can be observed that the estimated quantile contours coincide with the respective density contours of the bivariate normal distribution. The functional quantile curves are also estimated over time and are plotted in Fig. 1b along with the true mean function of the simulated data. The estimated quantile curves pick up the pattern of the true mean function well. The median functional quantile curve predicts the center of the functional distribution, while the curves at all the multivariate quantiles predict the complete functional distribution.

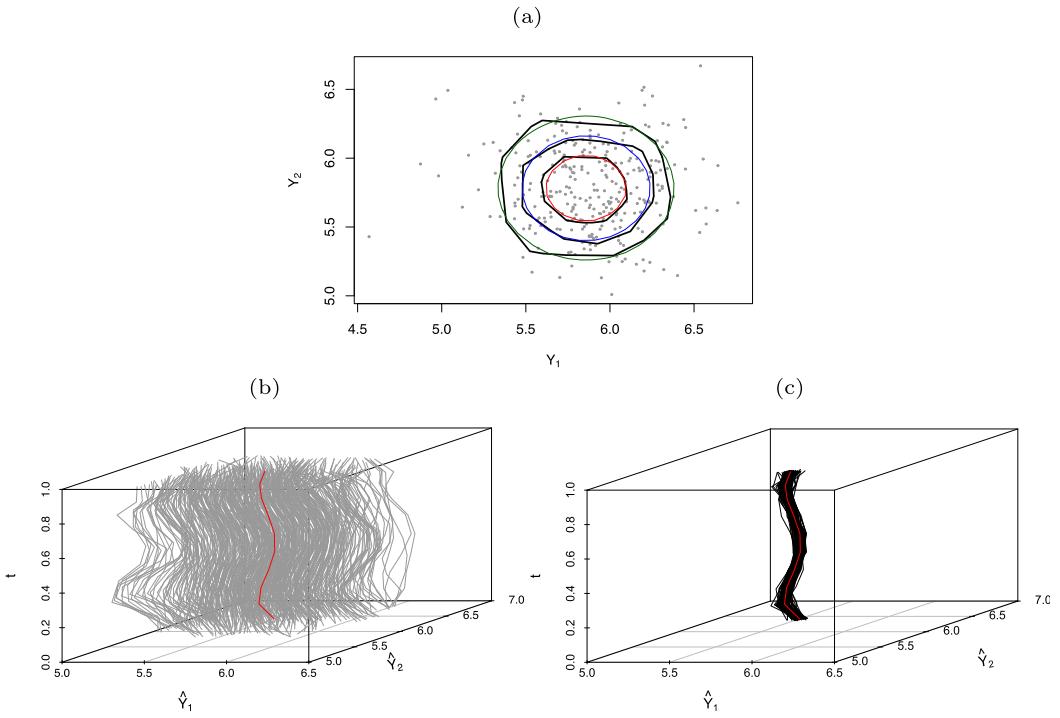

Fig. 1. (a) Predicted contours for simulated data at $t = 0.48$ for bivariate functional data with convex margins are in solid black for quantile levels 0.25, 0.5, and 0.75, with the density contours of the true bivariate normal distribution overlayed with colored lines for quantiles level 0.25 (red), 0.5 (blue), 0.75 (green); and (b) predicted functional quantile curves for the 441 points of the reference distribution $U(0,1)^2$ are shown in grey, and the true mean function is overlayed in red; (c) predicted median curves to estimate the center of the distribution for 100 simulations are in black, and the true center is in red. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

To assess the accuracy of the prediction model, we compare the predicted median quantile curve with the true mean function by repeating the simulation and computing the mean square prediction error (MSPE). The above simulation study is repeated 100 times, and the MSPE is calculated in the following way

$$\text{MSPE} = \frac{1}{100}\sum_{j=1}^{100} D_j(f(t), f_{\text{M}}(t)); \quad D_j(f(t), f_{\text{M}}(t)) = \frac{1}{p}\sum_{i=1}^{p} d(f(t_i), f_{\text{M}}(t_i)), \tag{7}$$

where d is the euclidean distance, and $f(t_i)$ and $f_{\text{M}}(t_i)$, $i = 1, \ldots, p$ are the true mean curves and predicted median curves, respectively. The obtained MSPE value for predicting the center of the simulated bivariate functional data is 0.1532. The predicted median quantile curves for all the simulations, plotted with the true mean function in Fig. 1c, are quite close to the true center of the distribution. They correctly pick up the pattern of the mean function over time. To deal with the uncertainty of the error distribution and estimation of contours, we average the contours over the 100 simulations. The averaged predicted contours for six quantile levels (0.1, 0.26, 0.42, 0.58, 0.74, 0.9) at $t = 0.48$ are plotted in Fig. 2a. The averaged contours are obtained by constructing the α-hull of the union of all sets of points of the predicted distribution from all the simulations.

We compare our method with the bivariate quantile varying coefficient model proposed by Pietrosanu et al. (2021). We estimate the directional quantile envelopes using Pietrosanu et al.'s method for the marginal distributions of the simulated data; the resulting quantile envelopes are plotted for $t = 0.48$, at six depth levels, in Fig. 2b. The resulting quantile envelopes are convex and work well in the case of the convex marginal distribution.

4.2. Bivariate functional data with nonconvex margins

In this simulation study, we simulate bivariate functional data $\boldsymbol{Y}(t) = (Y_1(t), Y_2(t))^\top$ with nonconvex error distribution using the following model

$$\begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1^\top \boldsymbol{\theta}(t) \\ \boldsymbol{X}_2^\top \boldsymbol{\theta}(t) \end{pmatrix} + \begin{pmatrix} Z + R\cos\phi \\ Z^2 + R\sin\phi \end{pmatrix},$$

where $Z \sim U[-0.5, 0.5]$, $\phi \sim U[0, 2\pi]$, $Z_0 \sim U[0, 0.5]$ are independent and $R = 0.2Z_0(0.5 + (0.5 - |Z|)/2)$. In the data generating model, $\boldsymbol{X}_1^\top \boldsymbol{\theta}(t)$ and $\boldsymbol{X}_2^\top \boldsymbol{\theta}(t)$ represents cubic spline functions in t with fixed number of knots. The setting of the

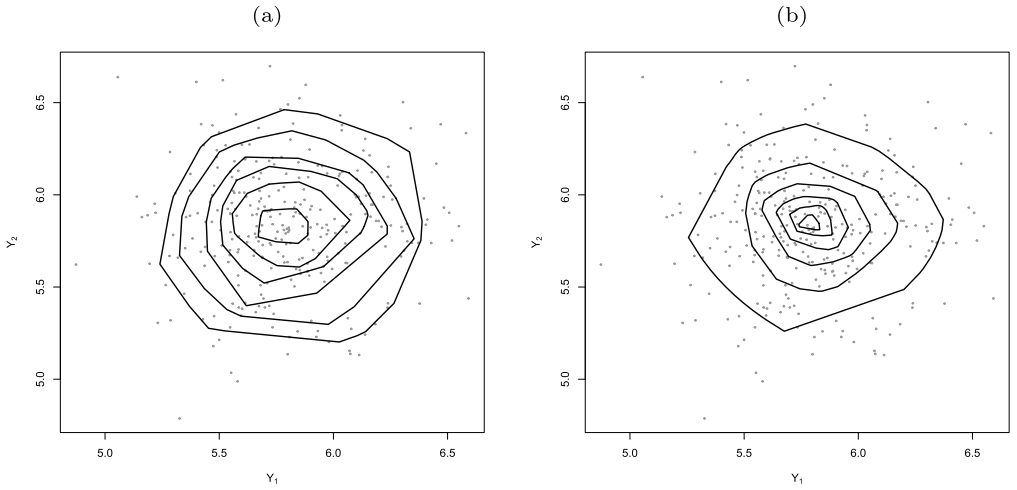

Fig. 2. Predicted contours for bivariate functional data with convex margins at $t = 0.48$ using (a) multivariate functional quantile model (2.1) in the paper for six quantile levels (0.1, 0.26, 0.42, 0.58, 0.74, 0.9) averaged over 100 simulations, and (b) Pietrosanu et al.'s method for six depth levels (0.01, 0.09, 0.17, 0.24, 0.32, 0.4).

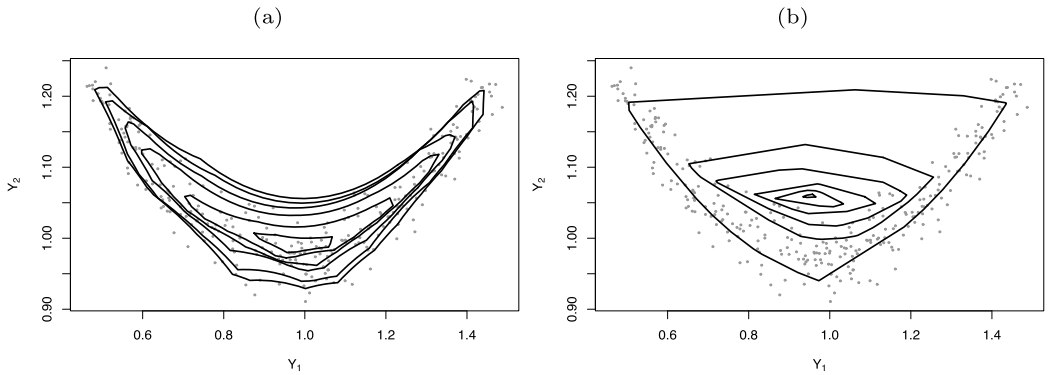

Fig. 3. Predicted contours for bivariate functional data with nonconvex margins at $t = 0.48$ using (a) multivariate functional quantile model (1) for six quantile levels (0.1, 0.26, 0.42, 0.58, 0.74, 0.9), and (b) Pietrosanu et al.'s method for six depth levels (0.01, 0.07, 0.13, 0.18, 0.24, 0.3).

mean function is similar to the previous simulation model. Data are generated for $m = 300$ and 10 equispaced points in [0,1].

Quantile contours are estimated for the nonconvex marginal distributions, and Fig. 3a shows the estimated quantile contours for the marginal distribution, at $t = 0.48$, for six quantile levels (0.1, 0.26, 0.42, 0.58, 0.74, 0.9). It can be seen that the estimated contours pick up the geometry of the distribution well, and are flexible with respect to the nonconvex shape of the distribution. We estimate the directional quantile envelopes using Pietrosanu et al.'s method, and the same is plotted for $t = 0.48$ at six depth levels in Fig. 3b. It can be observed that the quantile envelopes, which are essentially Tukey depth contours, cannot pick up the geometry of the distribution. This observation is common to most depth-based contours that follow the characteristics of convexity. For our method, the functional quantile curves are also estimated over time and plotted with the true mean function in Fig. 4. The pattern of the estimated quantile curves is quite similar to the true mean function.

4.3. Trivariate functional data with nonconvex margins

One of the advantages of the proposed multivariate functional quantile model is that it is computationally efficient for higher dimensions. In this section, we simulate trivariate functional data $\boldsymbol{Y}(t) = (Y_1(t), Y_2(t), Y_3(t))^\top$ with a nonconvex marginal distribution, using the following model

$$\begin{pmatrix} Y_1(t) \\ Y_2(t) \\ Y_3(t) \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1^\top \boldsymbol{\theta}(t) \\ \boldsymbol{X}_2^\top \boldsymbol{\theta}(t) \\ \boldsymbol{X}_3^\top \boldsymbol{\theta}(t) \end{pmatrix} + \begin{pmatrix} Z + R\cos\phi \\ Z^2 + R\sin\phi \\ Z + R \end{pmatrix},$$

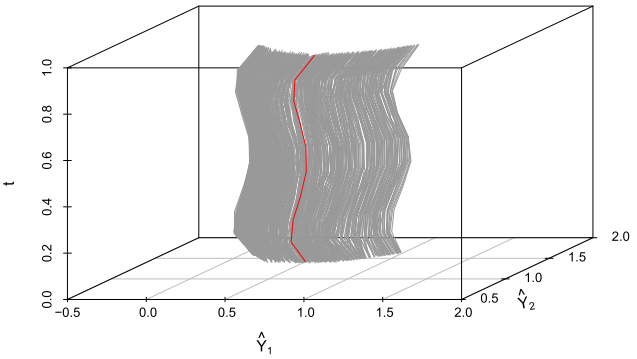

Fig. 4. Predicted functional quantile curves for bivariate functional data with nonconvex margins for the 441 points of the reference distribution $U(0,1)^2$ are shown in grey and the true mean function is overlayed in red.

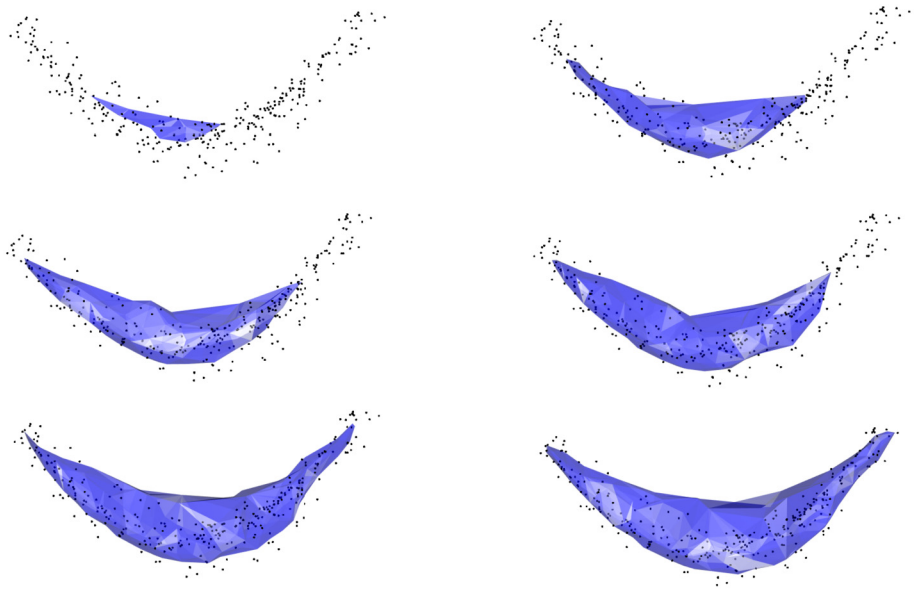

Fig. 5. Predicted quantiles shapes for trivariate functional data with nonconvex margins for six quantile levels (0.1, 0.26, 0.42, 0.58, 0.74, 0.9) at $t = 0.48$.

with similar notations as Section 4.2. We simulate data for $m = 300$, for 10 equispaced points in [0,1]. We use model (1) to estimate the multivariate quantile function. We choose the reference distribution to be $U(0,1)^3$. The reference distribution is generated on a $9 \times 9 \times 9$ uniformly spaced grid in with $n = 729$, and mapped to the spherical uniform reference distribution that was generated on spheres with evenly spaced radii in (0,1]. The multivariate quantiles are estimated for the marginal distributions of the simulated functional data. The quantile shapes are plotted to visualize the geometry of the trivariate distribution, at $t = 0.48$, for six quantile level, (0.1, 0.26, 0.42, 0.58, 0.74, 0.9) in Fig. 5. The quantile shapes can correctly obtain the geometry of the distribution in the trivariate case as well. To evaluate the performance of our method, we compare the center of the trivariate functional distribution, which is known, with the predicted center, using the predicted median quantile curve, and compute the MSPE, using (7). The simulation is repeated 100 times, and the obtained MSPE value is 0.2641.

5. Applications to air pollution data

We now illustrate the flexibility of our proposed methodology with an application to air pollution data. We consider air quality data across the Northeastern United States region (Karl and Koss, 1984) defined by the National Climatic Data Center (New Hampshire, Vermont, New York, Massachusetts, Connecticut, Rhode Island, Pennsylvania, New Jersey, Delaware) for the years 2011 and 2012. In particular, we study the joint distribution of $PM_{2.5}$ and geopotential height at 850 hPa over different months. The data for $PM_{2.5}$ is obtained from the Environmental Protection Agency (EPA), which provides a daily average of values generated from a Community Multiscale Air Quality Modeling System (CMAQ, https://www.epa.gov/cmaq). The data for geopotential height is obtained from the North American Regional Reanalysis (NARR, https://www.esrl.noaa.gov/psd), which provides monthly average values. We average the daily $PM_{2.5}$ concentrations with each month to

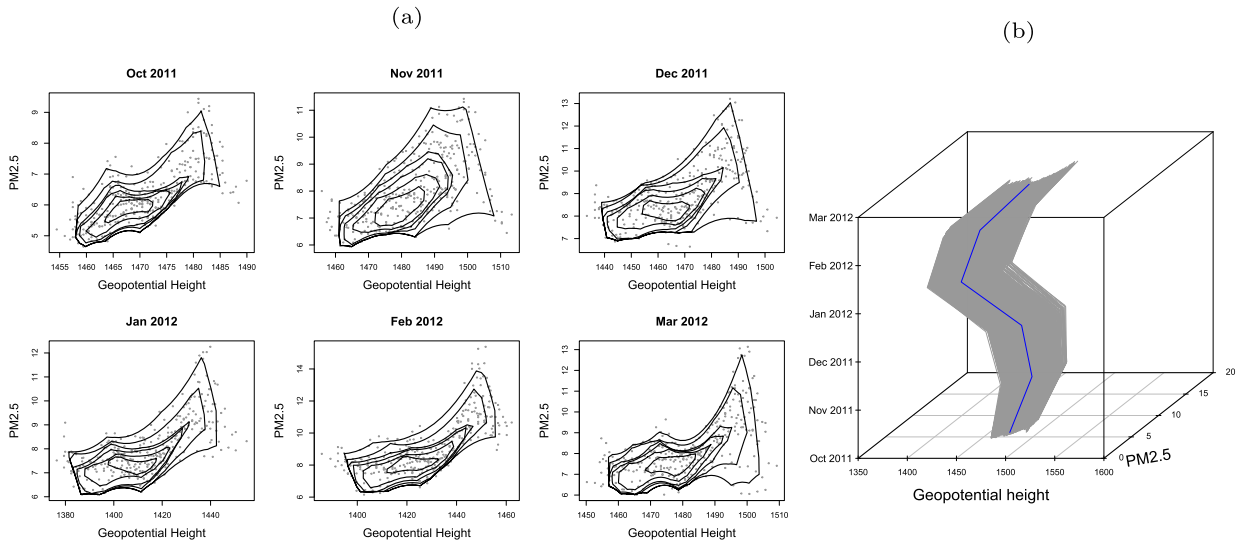

Fig. 6. (a) Predicted quantile contours for six quantile levels, (0.1, 0.26, 0.42, 0.58, 0.74, 0.9), and (b) predicted functional quantile curves over time for the distribution of $PM_{2.5}$ and geopotential height at 850 hPa from October 2011 to March 2012.

comply with monthly geopotential height data. Early work on estimating multivariate marginal distribution relied heavily on parametric distributional assumptions as pointed out by Wei and He (2006) or was restricted by the convexity of its contours (Pietrosanu et al., 2021), but recent research has emphasized the need for more flexible models to analyze such datasets (Chernozhukov et al., 2017).

Here, we apply the proposed methods to the bivariate distribution of the monthly averages of $PM_{2.5}$ and geopotential height at 850 hPa, across the various Northeastern states mentioned above, for two seasons, October 2011 to March 2012, and April 2012 to November 2012, and estimate the quantile contours for each month. We choose the widely used cubic B-splines in the proposed model with four knots as there are just six time points. The estimated quantile contours for October 2011 to March 2012 are shown in Fig. 6a. The distribution of $PM_{2.5}$ and geopotential height at 850 hPa for most of the months is nonconvex, and the contours estimate the geometry of the distribution reasonably well. The estimated marginal distribution also explains the variability of the distribution with the contours and provides valuable summary statistics such as the center of the distribution by estimating the bivariate median. From Fig. 6a, we can observe that in February 2012, the increase in the $PM_{2.5}$ concentration with the increase in geopotential height is not linear, and the median of the distribution can be seen as the change point in geopotential height from where the $PM_{2.5}$ concentration gets much higher. The results for the second season, April 2012 to November 2012, are plotted in Fig. 7. In the second season, many months have a bimodal distribution, which is clearly shown by the estimated contours, whereas most existing depth-based contours embodied with convexity cannot pick up such features.

Typically, univariate quantile curves are used to study the changes of a variable over time, but these methods aim at screening one variable at a time. Because $PM_{2.5}$ is significantly affected by geopotential height, the overall change should be assessed by studying bivariate quantile curves since univariate quantile curves fail to take correlations into account. The estimated bivariate functional quantile curves for the two seasons, October 2011 to March 2012, and April 2012 to November 2012, are shown in Figs. 6b and 7b, respectively. The functional quantile curves show the changes in the bivariate distribution of $PM_{2.5}$ and geopotential height, over time, for the whole distribution. We denote the quantile curve corresponding to the median of the reference distribution as the median quantile curve. The median curve summarizes the dynamic change in the distribution of $PM_{2.5}$ and geopotential height from October 2011 to March 2012. The median curves for the two seasons show different patterns of the bivariate distribution over time. The visualization of bivariate functional data is not trivial, and the estimated functional quantile curves visualize the centrality of curves and variability in the curves over the seasons, and the estimated contours visualize the geometry of the marginal distributions for each month.

6. Discussion

In this article, we propose a flexible model to estimate quantiles for multivariate functional data and develop the estimation and prediction procedures. The estimated quantile contours can account for non-Gaussian and even nonconvex multivariate distributions marginally without any parametric distributional assumptions. The estimated multivariate quantile function has desirable theoretical properties like monotonicity, uniqueness, and consistency. We also show that the method is computationally efficient for high-dimensional data (more than 2). The proposed methodology is demonstrated on $PM_{2.5}$ and geopotential height data over time across the Northeastern United States. The estimated contours pick up the flexible geometry of the marginal distributions, and the bivariate quantile curves capture the dynamic change over time.

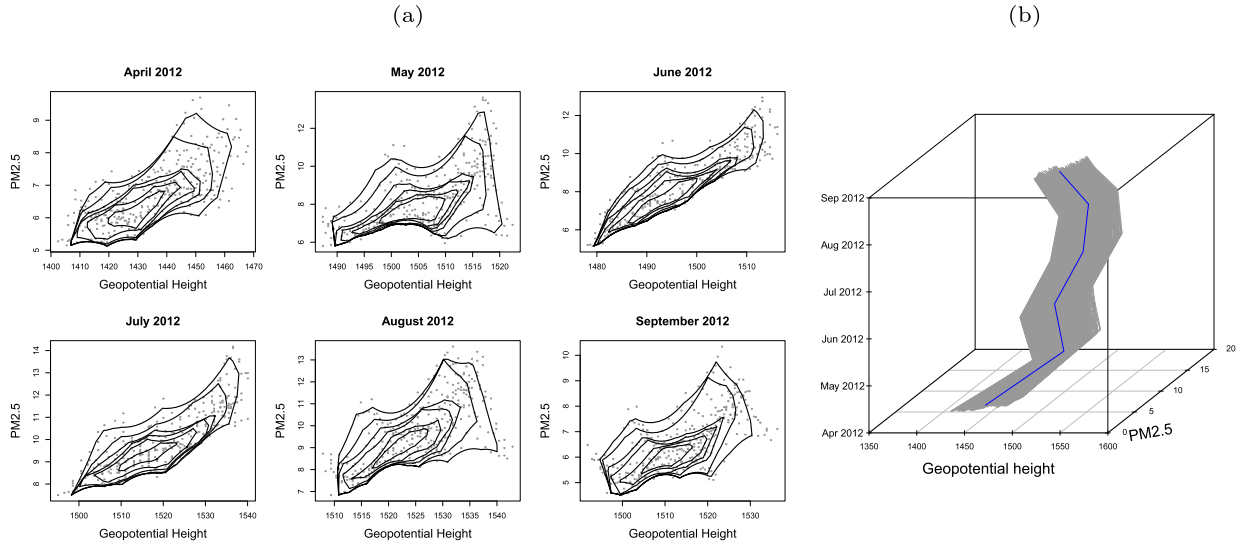

Fig. 7. (a) Predicted quantile contours for six quantile levels, (0.1, 0.26, 0.42, 0.58, 0.74, 0.9), and (b) predicted functional quantile curves over time for the distribution of $PM_{2.5}$ and geopotential height at 850 hPa from April 2012 to November 2012.

There still exist some interesting open problems that require future research. For example, we consider the intercept model for visualization in our simulation and application studies, but more covariates can be added in the model to study the dependence of the covariates on the response, using the multivariate quantiles. Moreover, in the cubic spline model for multivariate functional quantiles, we fix the number of knots. To choose the number of knots adaptively, we can use the cross-validation criterion by leaving out one curve at a time and use the estimated median curve as the prediction to compute the prediction errors. Then we choose the number of knots corresponding to the minimum mean prediction error.

Acknowledgements

This publication is based upon work supported by King Abdullah University of Science and Technology (KAUST), Office of Sponsored Research (OSR) under Award No: OSR-2019-CRG7-3800, and grant by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2018-04486).

Appendix A. Proof of Theorem 1

Proof. Consider the sequence $\hat{\boldsymbol{Q}}_n = \hat{\boldsymbol{Q}}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u}) = \nabla_{\boldsymbol{u}_n}\boldsymbol{B}(\boldsymbol{u}_n, t)^\top \boldsymbol{x}$, we want to prove that as $n \to \infty$, $\hat{\boldsymbol{Q}}_n$ converges uniformly to $\boldsymbol{Q} = \nabla_{\boldsymbol{u}}\boldsymbol{B}(\boldsymbol{u}, t)^\top \boldsymbol{x}$ for all $\boldsymbol{u} \in \mathcal{U}$, $\boldsymbol{x} \in \mathcal{X}$, and $t \in \mathcal{T}$.

Lemma 1. *Let $\mathbb{D}$ and $\mathbb{E}$ be compact sets and $g : \mathbb{D} \to \mathbb{E}$ is continuous. Then for any convergent sequence $x_n \to x$ in $\mathbb{D}$, the sequence of functions g_n converges uniformly to g if and only if $g_n(x_n) \to g(x)$.*

Lemma 1 is referred to as the extended continuous mapping theorem, for the proof see Theorem B.3 of Bücher et al. (2014). Using this lemma, if $g_n = g$, then $g_n \to g$ trivially and hence, $g(x_n) \to g(x)$ for any convergent sequence $x_n \to x$.

Hence, for any convergent sequence $\{\boldsymbol{u}_n\}$ such that $\boldsymbol{u}_n \to \boldsymbol{u}$, $\boldsymbol{u} \in \mathcal{U}$, using Lemma 1 we have

$$\boldsymbol{B}(\boldsymbol{u}_n, t)^\top \boldsymbol{x} \to \boldsymbol{B}(\boldsymbol{u}, t)^\top \boldsymbol{x}.$$

Lemma 2. *Consider a sequence $\{g_n\}$, such that $\lim_{n\to\infty} g_n(x_0)$ exists and is finite, $x_0 \in [a, b]$ and the sequence $\{\frac{\partial g_n}{\partial x}\}$ converges uniformly on $[a, b]$, then g_n uniformly converges to g and $\frac{\partial g_n}{\partial x}$ converges to $\frac{\partial g}{\partial x}$.*

For the proof of Lemma 2, see Theorem 7.17 of Rudin (1976). We denote $\nabla_{\boldsymbol{u}_n}\boldsymbol{B}(\boldsymbol{u}_n, t)^\top \boldsymbol{x} = \left(\frac{\partial}{\partial u_{1n}}\boldsymbol{B}(\boldsymbol{u}_n, t)^\top \boldsymbol{x}, \ldots, \frac{\partial}{\partial u_{dn}}\boldsymbol{B}(\boldsymbol{u}_n, t)^\top \boldsymbol{x}\right)$. Since $\nabla_{\boldsymbol{u}_n}\boldsymbol{B}(\boldsymbol{u}_n, t)^\top \boldsymbol{x}$ exists $F_{\boldsymbol{U}}$-almost everywhere on $\mathcal{U}$, the set $\bar{\mathcal{U}}$ where it does not exist is negligible. Therefore, the sequences $\frac{\partial}{\partial u_{1n}}\boldsymbol{B}(\boldsymbol{u}_n, t)^\top \boldsymbol{x}$, $\ldots$, $\frac{\partial}{\partial u_{dn}}\boldsymbol{B}(\boldsymbol{u}_n, t)^\top \boldsymbol{x}$ converges $F_{\boldsymbol{U}}$-almost everywhere on $\mathcal{U}$. Hence, using Lemma 2, we have

$$\frac{\partial}{\partial u_{1n}}\boldsymbol{B}(\boldsymbol{u}_n,t)^\top \boldsymbol{x} \to \frac{\partial}{\partial u_1}\boldsymbol{B}(\boldsymbol{u},t)^\top \boldsymbol{x}$$

$$\vdots$$

$$\frac{\partial}{\partial u_{dn}}\boldsymbol{B}(\boldsymbol{u}_n,t)^\top \boldsymbol{x} \to \frac{\partial}{\partial u_d}\boldsymbol{B}(\boldsymbol{u},t)^\top \boldsymbol{x}.$$

We have d sequences that converge uniformly; we now want to prove that they converge jointly. Consider d sequences $g_{1n},\dots g_{dn}$ that uniformly converge to $g_1,\dots,g_d$, respectively. From the definition of uniform convergence, we have the following:

For every $\epsilon_1 > 0$, there exists a natural number $N(\epsilon_1)$ such that $\mid g_{1n} - g_1 \mid < \epsilon_1$, for all $n \geq N(\epsilon_1)$. Similarly, for every $\epsilon_k > 0$, there exists $N(\epsilon_k)$ such that $|g_{dn} - g_d| < \epsilon_d$, for all $n \geq N(\epsilon_d)$.

Consider

$$\|(g_{1n},\dots,g_{dn}) - (g_1,\dots,g_d)\| = \sqrt{(g_{1n}-g_1)^2 + \cdots + (g_{dn}-g_d)^2}$$
$$\leq \sqrt{\epsilon_1^2 + \cdots + \epsilon_d^2} = \epsilon^*, \text{ for all } n \geq N = \max(N(\epsilon_1),\dots,N(\epsilon_d)).$$

For every $\epsilon^* > 0$, there exists a natural number $N = \max(N(\epsilon_1),\dots,N(\epsilon_d))$ such that

$$\|(g_{1n},\dots,g_{dn}) - (g_1,\dots,g_d)\| < \epsilon^*, \quad \text{for all } n > N.$$

Therefore,

$$(g_{1n},\dots,g_{dn}) \to (g_1,\dots,g_d).$$

Using this property of uniform convergence, we show that

$$\left(\frac{\partial}{\partial u_{1n}}\boldsymbol{B}(\boldsymbol{u}_n,t)^\top \boldsymbol{x},\dots,\frac{\partial}{\partial u_{dn}}\boldsymbol{B}(\boldsymbol{u}_n,t)^\top \boldsymbol{x}\right) \to \left(\frac{\partial}{\partial u_1}\boldsymbol{B}(\boldsymbol{u},t)^\top \boldsymbol{x},\dots,\frac{\partial}{\partial u_d}\boldsymbol{B}(\boldsymbol{u},t)^\top \boldsymbol{x}\right).$$

Hence, we conclude that

$$\hat{\boldsymbol{Q}}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u}) \to \boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u}).$$

This implies, as $n \to \infty$

$$\sup_{\boldsymbol{u}\in\mathcal{U}} \|\hat{\boldsymbol{Q}}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u}) - \boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u})\| \to 0,$$

i.e., $\hat{\boldsymbol{Q}}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u})$ uniformly converges to $\boldsymbol{Q}_{\boldsymbol{Y}(t)|\boldsymbol{X}}(\boldsymbol{u})$ for $u \in \mathcal{U}$, $F_{\boldsymbol{U}}$-almost everywhere. □

Appendix B. Proof of Corollary 1

Proof. The Hausdorff distance d_H between two non-empty subsets X and Y of a Euclidean metric space is given by

$$d_H(X,Y) = \max\{\sup_{x\in X}\inf_{y\in Y} d(x,y), \sup_{y\in Y}\inf_{x\in X} d(x,y)\},$$

where d is the euclidean distance. Let A be a non empty subset of $\mathcal{U}$, consider the Hausdorff distance between $\hat{\boldsymbol{Q}}(A)$ and $\boldsymbol{Q}(A)$,

$$\sup_{A\subseteq\mathcal{U}} d_H(\hat{\boldsymbol{Q}}(A), \boldsymbol{Q}(A)) = \sup_{A\subseteq\mathcal{U}}\left(\max\left\{\sup_{\boldsymbol{u}\in A}\inf_{\boldsymbol{u}^*\in A} d(\hat{\boldsymbol{Q}}(\boldsymbol{u}), \boldsymbol{Q}(\boldsymbol{u}^*)), \sup_{\boldsymbol{u}^*\in A}\inf_{\boldsymbol{u}\in A} d(\hat{\boldsymbol{Q}}(\boldsymbol{u}), \boldsymbol{Q}(\boldsymbol{u}^*))\right\}\right)$$
$$= \sup_{A\subseteq\mathcal{U}}\left(\max\left\{\sup_{\boldsymbol{u}\in A}\inf_{\boldsymbol{u}^*\in A} \|\hat{\boldsymbol{Q}}(\boldsymbol{u}) - \boldsymbol{Q}(\boldsymbol{u}^*)\|, \sup_{\boldsymbol{u}^*\in A}\inf_{\boldsymbol{u}\in A} \|\hat{\boldsymbol{Q}}(\boldsymbol{u}) - \boldsymbol{Q}(\boldsymbol{u}^*)\|\right\}\right)$$
$$\leq \sup_{A\subseteq\mathcal{U}}\left(\max\left\{\sup_{\boldsymbol{u}\in A} \|\hat{\boldsymbol{Q}}(\boldsymbol{u}) - \boldsymbol{Q}(\boldsymbol{u})\|, \sup_{\boldsymbol{u}^*\in A} \|\hat{\boldsymbol{Q}}(\boldsymbol{u}^*) - \boldsymbol{Q}(\boldsymbol{u}^*)\|\right\}\right)$$
$$= \sup_{A\subseteq\mathcal{U}}\left(\sup_{\boldsymbol{u}\in A} \|\hat{\boldsymbol{Q}}(\boldsymbol{u}) - \boldsymbol{Q}(\boldsymbol{u})\|\right)$$
$$= \sup_{\boldsymbol{u}\in\mathcal{U}} \|\hat{\boldsymbol{Q}}(\boldsymbol{u}) - \boldsymbol{Q}(\boldsymbol{u})\| \to 0, \quad \text{as } n \to \infty \quad \text{(using Theorem 1)}.$$

This implies, as $n \to \infty$

$$\sup_{A \subseteq \mathcal{U}} d_H(\hat{\boldsymbol{Q}}(A), \boldsymbol{Q}(A)) \to 0.$$

Hence, proved. □

Appendix C. Technical details

The dual problem (6) was obtained from the primal problem (5) using Lagrange multipliers, as follows:

$\max_{\text{vec(p)}\geq 0} \text{vec}\left(\text{UY}^\top\right)^\top \text{vec(p)}$ is equivalent to $\min_{\text{vec(p)}\geq 0} -\text{vec}\left(\text{UY}^\top\right)^\top \text{vec(p)}$ with same constraints. Consider the Lagrangian function,

$$L(\text{p}, \psi, \text{b}, \lambda) = -\text{vec}\left(\text{UY}^\top\right)^\top \text{vec(p)} + \text{vec}(\psi)^\top \left(\left(I_{mr} \otimes 1_n^\top\right)\text{vec(p)} - \nu\right) + \text{vec(b)}^\top \left(\left(\text{X}^\top \otimes I_n\right)\text{vec(p)} - \text{vec}\left(\mu\nu^\top \text{X}\right)\right) - \lambda^\top \text{vec(p)}$$

Taking partial derivative of the Lagrangian function with respect to vec(p) and equating it to zero

$$\frac{\partial L(\text{p}, \psi, \text{b}, \lambda)}{\partial \text{vec(p)}} = -\text{vec}\left(\text{UY}^\top\right)^\top + \left(I_{mr} \otimes 1_n\right)\psi + \left(\text{X} \otimes I_n\right)\text{vec(b)} - \lambda = 0$$

Using dual feasibility: $\lambda \geq 0$, we have

$$\left(\text{X} \otimes I_n\right)\text{vec(b)} - \text{vec}\left(\text{UY}^\top\right)^\top + \left(I_{mr} \otimes 1_n\right)\psi \geq 0. \tag{C.1}$$

Rearranging the Lagrangian function we get,

$$L(\text{p}, \psi, \text{b}, \lambda) = -\psi^\top \nu - \text{vec(b)}^\top \text{vec}(\mu\nu^\top \text{X}) + \left(-\text{vec}\left(\text{UY}^\top\right)^\top + \text{vec}(\psi)^\top \left(I_{mr} \otimes 1_n^\top\right) + \text{vec(b)}^\top \left(\text{X}^\top \otimes I_n\right)\right)\text{vec(p)}.$$

The terms with vec(p) are neglected and the dual problem reduces to the minimization of $-\psi^\top \nu - \text{vec(b)}^\top \text{vec}(\mu\nu^\top \text{X})$ subject to constraint $\lambda \geq 0$ which is equivalent to maximizing $\psi^\top \nu + \text{vec(b)}^\top \text{vec}(\mu\nu^\top \text{X})$ subject to constraint (C.1). Hence the dual problem is

$$\min_{\psi, \text{b}} \psi^\top \nu + \text{vec(b)}^\top \text{vec}(\mu\nu^\top \text{X}) : \quad \left(I_{mr} \otimes 1_n\right)\psi + \left(\text{X} \otimes I_n\right)\text{vec(b)} \geq \text{vec}(\text{UY}^\top),$$

where ψ is $mr \times 1$ vector and b is a $n \times K$ matrix.

Appendix D. Supplementary material

The supplementary material includes the R code used in this research and links to the datasets used in the application. It has been made publicly available on https://github.com/agarwag/Flexible-quantile-contours-multivariate-functional-data.

References

Agarwal, G., Sun, Y., 2020. Bivariate functional quantile envelopes with application to radiosonde wind data. Technometrics, 1–13.

Breckling, J., Kokic, P., Lübke, O., 2001. A note on multivariate m-quantiles. Stat. Probab. Lett. 55 (1), 39–44.

Brenier, Y., 1991. Polar factorization and monotone rearrangement of vector-valued functions. Commun. Pure Appl. Math. 44 (4), 375–417.

Bücher, A., Segers, J., Volgushev, S., 2014. When uniform weak convergence fails: empirical processes for dependence functions and residuals via epi- and hypographs. Ann. Stat. 42 (4), 1598–1634.

Burkard, R.E., Dell'Amico, M., Martello, S., 2009. Assignment Problems. Springer.

Carlier, G., Chernozhukov, V., Galichon, A., 2016. Vector quantile regression: an optimal transport approach. Ann. Stat. 44 (3), 1165–1192.

Chakraborty, A., Chaudhuri, P., 2014. The spatial distribution in infinite dimensional spaces and related quantiles and depths. Ann. Stat. 42 (3), 1203–1231.

Chakraborty, B., 2001. On affine equivariant multivariate quantiles. Ann. Inst. Stat. Math. 53 (2), 380–403.

Chakraborty, B., 2003. On multivariate quantile regression. J. Stat. Plan. Inference 110 (1), 109–132.

Chaudhuri, P., 1996. On a geometric notion of quantiles for multivariate data. J. Am. Stat. Assoc. 91 (434), 862–872.

Chernozhukov, V., Galichon, A., Hallin, M., Henry, M., 2017. Monge–Kantorovich depth, quantiles, ranks and signs. Ann. Stat. 45 (1), 223–256.

Chowdhury, J., Chaudhuri, P., 2017. Nonparametric Quantile Regression for Banach-Valued Response. CRC Press, pp. 225–251, Chapter 14.

Chowdhury, J., Chaudhuri, P., 2019. Nonparametric depth and quantile regression for functional data. Bernoulli 25 (1), 395–423.

Cohen, A.J., Ross Anderson, H., Ostro, B., Pandey, K.D., Krzyzanowski, M., Künzli, N., Gutschmidt, K., Pope, A., Romieu, I., Samet, J.M., 2005. The global burden of disease due to outdoor air pollution. J. Toxicol. Environ. Health, Part A 68 (13–14), 1301–1307.
Edelsbrunner, H., Kirkpatrick, D., Seidel, R., 1983. On the shape of a set of points in the plane. IEEE Trans. Inf. Theory 29 (4), 551–559.
Fraiman, R., Pateiro-López, B., 2012. Quantiles for finite and infinite dimensional data. J. Multivar. Anal. 108, 1–14.
Fuentes, M., Song, H.-R., Ghosh, S.K., Holland, D.M., Davis, J.M., 2006. Spatial association between speciated fine particles and mortality. Biometrics 62 (3), 855–863.
Hallin, M., Paindaveine, D., Siman, M., 2010. Multivariate quantiles and multiple-output regression quantiles: from l_1 optimization to halfspace depth. Ann. Stat. 38 (2), 635–669.
He, Q., Geng, F., Li, C., Mu, H., Zhou, G., Liu, X., Gao, W., Wang, Y., Cheng, T., 2018. Long-term variation of satellite-based pm 2.5 and influence factors over east China. Sci. Rep. 8 (1), 1–10.
Hettmansperger, T., Nyblom, J., Oja, H., 1992. On multivariate notions of sign and rank. In: Dodge, Y. (Ed.), L1-Statistical Analysis and Related Methods. North Holland, Amsterdam, pp. 267–278.
Hu, Z., Liebens, J., Rao, K.R., 2008. Linking stroke mortality with air pollution, income, and greenness in northwest Florida: an ecological geographical study. Int. J. Health Geogr. 7 (1), 20.
Huang, J.Z., Wu, C.O., Zhou, L., 2002. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. Biometrika 89 (1), 111–128.
Karl, T., Koss, W.J., 1984. Regional and National Monthly, Seasonal, and Annual Temperature Weighted by Area. National Climatic Data Center (U.S.), pp. 1895–1983.
Kong, L., Mizera, I., 2012. Quantile tomography: using quantiles with multivariate data. Stat. Sin. 22 (4), 1589–1610.
Li, Y., Chen, Q., Zhao, H., Wang, L., Tao, R., 2015. Variations in pm10, pm2.5 and pm1.0 in an urban area of the Sichuan basin and their relation to meteorological factors. Atmosphere 6 (1), 150–163.
Liu, R.Y., 1990. On a notion of data depth based on random simplices. Ann. Stat., 405–414.
Liu, X., Zuo, Y., Wang, Z., 2013. Exactly computing bivariate projection depth contours and median. Comput. Stat. Data Anal. 60 (1–11), 0167.
López-Pintado, S., Romo, J., 2009. On the concept of depth for functional data. J. Am. Stat. Assoc. 104 (486), 718–734.
McCann, R.J., 1995. Existence and uniqueness of monotone measure-preserving maps. Duke Math. J. 80 (2), 309–324.
Minty, G.J., 1964. On the monotonicity of the gradient of a convex function. Pac. J. Math. 14 (1), 243–247.
Murty, K.G., 1983. Linear Programming. Wiley.
Paindaveine, D., Šiman, M., 2012. Computing multiple-output regression quantile regions. Comput. Stat. Data Anal. (ISSN 0167-9473) 56 (4), 840–853.
Papadimitriou, C.H., Steiglitz, K., 1982. Combinatorial Optimization: Algorithms and Complexity. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
Pietrosanu, M., Shu, H., Jiang, B., Kong, L., Heo, G., He, Q., Gilmore, J., Zhu, H., 2021. Estimation for the bivariate quantile varying coefficient model with application to diffusion tensor imaging data analysis. Biostatistics. https://academic.oup.com/biostatistics/advance-article-abstract/doi/10.1093/biostatistics/kxab031/6356199.
Rudin, W., 1976. Principles of Mathematical Analysis. McGraw-Hill.
Russell, B.T., Wang, D., McMahan, C.S., 2017. Spatially modeling the effects of meteorological drivers of pm2.5 in the eastern United States via a local linear penalized quantile regression estimator. Environmetrics 28 (5), e2448.
Saunders, R.O., Waugh, D.W., 2015. Variability and potential sources of summer pm2.5 in the northeastern United States. Atmos. Environ. 117, 259–270.
Serfling, R., 2010. Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardisation. J. Nonparametr. Stat. 22 (7), 915–936.
Serfling, R., Wijesuriya, U., 2017. Depth-based nonparametric description of functional data, with emphasis on use of spatial depth. Comput. Stat. Data Anal. 105, 24–45.
Sun, Y., Genton, M.G., 2011. Functional boxplots. J. Comput. Graph. Stat. 20 (2), 316–334.
Tukey, J.W., 1975. Mathematics and the picturing of data. In: Proceedings of the International Congress of Mathematicians, vol. 2, pp. 523–531.
Villani, C., 2003. Topics in Optimal Transportation, vol. 58. American Mathematical Soc.
Wei, Y., 2008. An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts. J. Am. Stat. Assoc. 103 (481), 397–409.
Wei, Y., He, X., 2006. Conditional growth charts. Ann. Stat. 34 (5), 2069–2097.
World Health Organization, 2003. Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide: report on a WHO working group, Bonn, Germany, 13-15 January 2003. Technical Report EUR/03/5042688. WHO Regional Office for Europe, Copenhagen.
Zuo, Y., 2019. A new approach for the computation of halfspace depth in high dimensions. Commun. Stat., Simul. Comput. 48 (3), 900–921.
Zuo, Y., Serfling, R., 2000. General notions of statistical depth function. Ann. Stat. 28 (2), 461–482.