



Robust functional multivariate analysis of variance with environmental applications

Zhuo Qu¹ | Wenlin Dai² | Marc G. Genton¹

¹Statistics Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

²Institute of Statistics and Big Data, Renmin University of China, Beijing, China

Correspondence

Wenlin Dai, Institute of Statistics and Big Data, Renmin University of China, Beijing 100086, China
Email: wenlin.dai@ruc.edu.cn

Funding information

King Abdullah University of Science and Technology and National Natural Science Foundation of China

Abstract

We propose median polish for functional multivariate analysis of variance (FMANOVA) with the implementation of depth for multivariate functional data. As an alternative to classical mean estimation, functional median polish estimates the functional grand effect and factor effects based on functional medians in one-way and two-way additive FMANOVA models. Median polish estimates in FMANOVA are visually unbiased, independently of the choice of multivariate functional depth. The corresponding mean-based and rank-based tests are generalized to evaluate whether the functional medians in various levels of the factors are the same. Simulation studies illustrate the robustness of our functional median polish in various scenarios, compared with the results from classical FMANOVA fitted by means. The results are evaluated both marginally and jointly. Three environmental data sets are considered to illustrate that our median polish is robust against outliers in practical implementations. Functional boxplots and heat maps are two ways of visualizing the functional factors, depending on whether the functional data are curves or images, respectively.

KEY WORDS

functional data, image data, median polish, multivariate analysis of variance, robustness, spatiotemporal data

1 | INTRODUCTION

Analysis of variance (ANOVA) is essential for studying the effects of categorical factors on a response in additive models. Depending on the number of factors, we can distinguish between one-way and multiway ANOVAs. When there is only one factor α to split the data group, we adopt a one-way ANOVA to compare the means of the groups in different categorical levels. When there are two factors, α and β , to divide the data group, we use a two-way ANOVA. The two-way ANOVA may have an interaction term ($(\alpha\beta)$, Kutner, Nachtsheim, Neter, & Li, 2005, chapter 19.2), which means that, for factors α and β , the effects of the categorical levels of α on the dependent variables are not consistent among the different categorical levels of β , or vice versa. If the responses in additive models consist of at least two dependent variables, and if these variables are moderately correlated, we can use multivariate analysis of variance (MANOVA, Bray & Maxwell, 1985) to study the effects of categorical factors on the responses. As a generalization of ANOVA, MANOVA examines the impacts of categorical factors on several dependent variables. By decomposing the variance-covariance matrix of dependent variables among the different factors, factors making significant contributions to the responses can be identified. How much of their variability is ascribable to each factor can also be determined.

Both ANOVA and MANOVA models can estimate the grand/factor effects with overall/groupwise means, which are sensitive to abnormal observations. For MANOVA, the assumption is that no joint outliers exist in the same level group. Compared with marginal outliers, joint outliers have an unusual combination of dependent variables, meaning that they are not necessarily univariate outliers, if seen marginally. We can extend ANOVA to functional ANOVA (FANOVA, Kaufman & Sain, 2010) and MANOVA to functional MANOVA (FMANOVA, Górecki & Smaga, 2017) by turning each response to a function of time or locations. Similarly to ANOVA (MANOVA), FANOVA (FMANOVA) assumes that there are no outliers in each group.

To address the potential outliers, Tukey (1970) and Mosteller and Tukey (1977) proposed the median polish for two-way additive models without an interaction term. As an alternative to classical ANOVA, median polish is a robust technique for studying the effects of various factors on the response, in an additive model. It replaces mean estimation with the median to identify all factors, hence is under little influence from outliers (Emerson & Hoaglin, 1983).

A two-way median polish demands iterations for the convergence of functional effects, which reduces the residuals in L_1 norm in the iteration process. Siegel (1983) proved that the iteration process in the median polish must stop if the table entries are rational numbers, but can take arbitrarily long, even for integer tables of a given fixed size. He then focused on the low median polish in a general matrix and proved that it must converge after a few iterations. The low median is defined to be the median if the sample size is an odd integer, and the lower of the middle two values if the sample size is even. Fink (1988) stated that for commensurable data, the algorithm converges in a finite number of steps, while it does not, in general, converge to the least L_1 norm residual. Furthermore, he proposed an algorithm that converges after finite iterations for any real data and provided the least residuals in L_1 norm. Chen and Farnsworth (1990) discussed that, for most small dimensions (i.e., one factor effect no greater than three or both factor effects equal three), median polish converges in no more than two iterations. Therefore, we expect that median polish converges in a small number of iterations practically.

In addition, median polish has many applications in spatial statistics and functional data. For instance, Cressie (1984) proposed the median polish kriging as a hybrid method of predicting spatial data, which unites kriging and analysis of tables. One can then use residuals of median polish to construct the stationary covariance function. Berke (2001) applied median polish in universal kriging for spatial data as a robust spatial prediction method to cover the problem of both the spatial trend effects and possible outliers. Sun and Genton (2012b) generalized median polish to the univariate functional data domain. Here, by building one-way and two-way FMANOVA models and by analyzing appropriate notions of multivariate functional depth, we develop a robust methodology to estimate the functional grand effect and factor effects with functional median polish. We only consider grand and additive factor effects in the two-way model for multivariate functional data, since the median polish assumes no interaction among factors in an additive ANOVA model.

Medians for multivariate data are usually defined via the notion of *statistical depth*. It provides an order from the center outward and can be used to find the “center” of a sample of discretely observed curves with the highest depth value and screen for outliers with small depths, for example, halfspace depth (HD, Tukey, 1975), simplicial depth (SD, Liu, 1990), projection-based depth (PD, Zuo, 2003), random Tukey depth (RTD, Cuesta-Albertos & Nieto-Reyes, 2008), skewness-adjusted projection depth (SPD, Hubert & Van der Veeken, 2008), and directional projection depth (DPD, Rousseeuw, Raymaekers, & Hubert, 2018). The depth for multivariate data can be extended to univariate functional data, for example, integrated depth (ID, Fraiman & Muniz, 2001), h-mode depth, random projection depth, and double random projection depth (hMD, RPD, RP2D, Cuevas, Febrero, & Fraiman, 2007), and band depth and modified band depth (BD, MBD, López-Pintado & Romo, 2009). With the algorithm developed by Sun, Genton, and Nychka (2012), the band depth of large functional data sets can be computed efficiently.

Correspondingly, several generalizations of depths have been proposed for multivariate functional data. Claeskens, Hubert, Slaets, and Vakili (2014) defined multivariate functional depth (MFD) from the multivariate depth at each time point and introduced a weight function to consider the weighted average at all time points, where halfspace depth is used as a basis depth specifically to define multivariate functional halfspace depth (MFHD). Combining the concept of MFD and multivariate depth mentioned above, we can define various multivariate functional depths, such as multivariate functional projection-based depth (MFPD), multivariate functional skewness-adjusted projection depth (MFSPD), multivariate functional directional projection depth (MFDPD), and multivariate functional simplicial depth (MFSD). Ieva and Paganoni (2013) introduced a multivariate band depth by taking every univariate component of the multivariate data into account. López-Pintado, Sun, Lin, and Genton (2014) generalized univariate functional band depth (BD, MBD) to simplicial band depth (SBD, MSBD) and provided natural criteria to order multivariate functional curves. SBD has the advantage of detecting outliers with a higher probability over marginal MBD and WMBD proposed by Ieva and Paganoni (2013). It

should be noted that MSBD can also be defined as putting simplicial depth (SD) in the multivariate functional depth, so MSBD and MFSD are equivalent. We use MSBD to refer to this depth later in the article. Dai and Genton (2019) introduced a measure of directional outlyingness (DO) that is suitable for both univariate and multivariate functional data. Specifically, they adopted robust Mahalanobis distance (RMD) to measure the relative distance of outlyingness to the median.

The introduction of depth to functional data makes ranking observations, finding the median, performing robust estimation, and detecting outliers possible. Sun and Genton (2011, 2012a) proposed functional boxplots and adjusted functional boxplots to visualize functional data based on band depth (BD) and modified band depth (MBD), although other rankings can be used. Genton, Johnson, Potter, Stenchikov, and Sun (2014) extended this visualization tool to surface boxplots. Sun and Genton (2012b) applied functional median polish to one-way and two-way FANOVA and determined the robustness of median polish compared with the mean version. With rank-based depths and distance-based depths, Dai and Genton (2018, 2019) introduced directional outlyingness to detect magnitude outliers, shape outliers, and their combination in multivariate functional data.

This article is organized as follows. The description of one-way and two-way additive FANOVA, the algorithms for median polish estimation, the notion of depth for multivariate functional data, and the hypothesis tests in FANOVA are all presented in Section 2. Monte Carlo simulation studies for different models and contrasts between robust median estimation and classical estimation by means are discussed in Section 3. The proposed methods assessed on three practical data sets from environmental science are presented in Section 4, along with the corresponding hypothesis tests. Further discussions are provided in Section 5.

To demonstrate the generality of median polish for multivariate functional data, we adopt five multivariate functional depths (MFDs) based on various depths for multivariate data: multivariate functional halfspace depth (MFHD), multivariate functional projection-based depth (MFPD), multivariate functional skewness-adjusted projection depth (MFSPD), multivariate functional directional projection depth (MFDPD), and multivariate functional simplicial depth [MFSD, which is exactly modified simplicial band depth (MSBD)]. In addition, we use robust Mahalanobis distance (RMD) for multivariate functional data and modified band depth (MBD) for treating multivariate functional data as several univariate functional data independently for each variable. Among the first six depths (or distance) in the multivariate functional perspective, five multivariate functional depths (MFD) gave similar estimation results; additionally, RMD was not good at detecting the median and always gave a larger estimation range. Therefore, we illustrate the robustness of our method with the choice of modified simplicial band depth (MSBD) in the simulations of Section 3 and show the estimations from median polish based on marginal MBD and from means as well. The simulation results with other depths are shown in the Supplementary Material. In the applications in Section 4, we keep MSBD as representative of multivariate functional depth and show the estimation results of median polish using MSBD and those of FANOVA fitted by means.

2 | FUNCTIONAL MULTIVARIATE ANALYSIS OF VARIANCE ALGORITHM

2.1 | One-way FANOVA algorithm with median polish

Suppose that we are interested in a functional additive model with only one factor. After obtaining the multivariate functional data observations for p -dependent variables at all levels, we want to estimate the functional grand effect and the functional factor effects (here we call them row effects). Each observation can be decomposed using a one-way functional additive model:

$$\mathbf{Y}_{i,k}(t) = \boldsymbol{\mu}(t) + \boldsymbol{\alpha}_i(t) + \boldsymbol{\epsilon}_{i,k}(t), \quad i = 1, \dots, r, \quad k = 1, \dots, m_i, \quad t \in \mathfrak{F}, \quad (1)$$

where $\mathbf{Y}_{i,k}(t) = (Y_{i,1,k}(t), \dots, Y_{i,p,k}(t))^\top$, $\boldsymbol{\mu}(t) = (\mu_1(t), \dots, \mu_p(t))^\top$, $\boldsymbol{\alpha}_i(t) = (\alpha_{i,1}(t), \dots, \alpha_{i,p}(t))^\top$, and $\boldsymbol{\epsilon}_{i,k}(t) = (\epsilon_{i,1,k}(t), \dots, \epsilon_{i,p,k}(t))^\top$. Here $r \geq 2$ is the number of rows, and m_i is the number of observations in the i th row. Also $\boldsymbol{\mu}(t)$ is a functional grand effect, $\boldsymbol{\alpha}_i(t)$ is the i th functional row effect, and $\boldsymbol{\epsilon}_{i,k}(t)$ is the k th measurement error in the i th row, with constraints, $\forall t \in \mathfrak{F}$, that $\text{median}_i\{\boldsymbol{\alpha}_i(t)\} = \mathbf{0}$ and $\text{median}_i\{\boldsymbol{\epsilon}_{i,k}(t)\} = \mathbf{0}$ for all k . When t denotes time, \mathfrak{F} is a region in \mathbb{R} . When t denotes location, we consider that \mathfrak{F} is a region in \mathbb{R}^2 . In addition, \mathfrak{F} can be a region in \mathbb{R}^3 for three-dimensional locations and a region in $\mathbb{R}^2 \times \mathbb{R}$ for space-time data. Then, we develop the following one-way median polish algorithm (see Algorithm 1 for more details).

Algorithm 1. One-way FMANOVA with median polish

1. Divide the multivariate functional data according to different levels of the row effect.
 2. According to multivariate functional depth, find the deepest observation in each row as the functional row median. Update the data by subtracting the functional row median from each functional data in that row.
 3. Calculate the functional median of the functional row medians, and set it as a functional grand effect $\mu(t)$.
 4. Subtract the functional grand effect from functional row medians and set them as functional row effects $\alpha_i(t)$.
-

2.2 | Two-way FMANOVA algorithm with median polish

Suppose that we are interested in a functional additive model with two factors. After obtaining the multivariate functional data observations for p -dependent variables, at all levels, for those two factors, we want to estimate the functional grand effect and the functional factor effects (here we call them row effects and column effects). Each observation can be decomposed using a two-way functional additive model:

$$\mathbf{Y}_{i,j,k}(t) = \mu(t) + \alpha_i(t) + \beta_j(t) + \epsilon_{i,j,k}(t), \quad i = 1, \dots, r, \quad j = 1, \dots, c, \quad k = 1, \dots, m_{ij}, \quad t \in \mathfrak{F}, \quad (2)$$

where $\mu(t)$, $\alpha_i(t)$ and r and \mathfrak{F} are the same as those in model (1) in Section 2.1, plus $\mathbf{Y}_{i,j,k}(t) = (Y_{i,j,1,k}(t), \dots, Y_{i,j,p,k}(t))^\top$, $\beta_j(t) = (\beta_{j,1}(t), \dots, \beta_{j,p}(t))^\top$, $\epsilon_{i,j,k}(t) = (\epsilon_{i,j,1,k}(t), \dots, \epsilon_{i,j,p,k}(t))^\top$. Here, $c \geq 2$ is the number of columns, m_{ij} is the number of observations in the i th row, and j th column cell, $\beta_j(t)$ is the j th functional column effect, and $\epsilon_{i,j,k}(t)$ is the k th measurement error in the cell of the i th row and the j th column. Moreover, we assume $\forall t \in \mathfrak{F}$, $\text{median}_i\{\alpha_i(t)\} = \text{median}_j\{\beta_j(t)\} = \mathbf{0}$ and $\text{median}_i\{\epsilon_{i,j,k}(t)\} = \text{median}_j\{\epsilon_{i,j,k}(t)\} = \mathbf{0}$ for all k . then, we propose the following two-way median polish algorithm (Emerson & Hoaglin, 1983; Mosteller & Tukey, 1977; see Algorithm 2 for more details).

Algorithm 2. Two-way FMANOVA with median polish

1. Divide the multivariate functional data according to different levels of row and column effects. We obtain the row effects and the column effects sequentially.
 2. According to the multivariate functional depth, choose the deepest observation in each row as the functional row median. Update the data by subtracting the functional row median from each multivariate functional data in that row.
 3. Calculate the functional median of the functional row medians, and set it as a functional grand effect $\mu(t)$. Subtract this functional median from functional row medians, and set them as functional row effects $\alpha_i(t)$.
 4. Then choose the deepest observation in each column according to the multivariate functional depth as the functional column median. Update the data by subtracting the functional column median from each multivariate functional data in that column.
 5. Calculate the functional median of the functional column medians and update the grand effect $\mu(t)$ by adding that median. Subtract the functional median from functional column medians and set them as functional column effects $\beta_j(t)$.
 6. Repeat Steps 2–5 and update the functional grand effect $\mu(t)$, functional row effects $\alpha_i(t)$, and column effects $\beta_j(t)$ at each loop until no change (or negligible change) occur with functional row effects.
-

The median is unique when the data size is odd, and is an average when the data size is even. There are cases in which the median needs to be chosen clearly, as shown below:

1. In Step 2 of Algorithm 1 (Steps 2 and 4 of Algorithm 2), when the median is not uniquely defined, we take the median as the average of functional curves in this situation.
2. In Step 3 of Algorithm 1 (Steps 3 and 5 of Algorithm 2), given only two row (or column) levels, we take the median as the average of two functional curves.

We extend the median polish to multivariate functional data from Tukey's median polish, which was originally for smoothing data in two-way tables. The convergence of median polish is discussed in the L_1 norm of the residual matrix $\mathbf{R} = (r_{ij})$. Although there is no proof of convergence of median polish for a general matrix, Siegel (1983) proved

convergence of two-way median polish for tables with rational numbers as entries. Chen and Farnsworth (1990) discussed several cases in median polish about convergence speed: for a two-way table with dimensions $r \times c$, if $r \geq 3$ and $c > 3$ (and vice versa), it is possible for the median polish to take an arbitrarily large or infinite number of steps to converge; if $r = c = 3$ or $r = 2, c > 2$ (and vice versa), median polish stops in no more than three half iterations, where each row or column polish is called a half iteration (Siegel, 1983).

To consider the convergence of median polish for multivariate functional data, we hereby consider the convergence in L_1 norm of the residuals r_{ij} . In this case, each entry is the sum of residuals at various times for all variables: $r_{ij} = \sum_{l=1}^p \sum_k R_{ij}^{(l)}(t_k)$, where the residuals in one cell are $\mathbf{R}_{ij}(t_k) = (R_{ij}^{(1)}(t_k), \dots, R_{ij}^{(p)}(t_k))^T$ and $t_k \in \mathfrak{F}$. Since the algorithm applied to functional data only changes the data structure and the method of obtaining the median, the convergence in L_1 norm is not expected to be affected. Specifically, when one category level is two and another one is greater than two, or both category levels are three, median polish stops in no more than three half iterations.

Similarly, when the median is not uniquely or strictly defined, taking the median as the average of samples does not influence the convergence of the algorithm. Acknowledging the existence of arbitrarily large numbers of iterations before convergence for cases in which $r > 3$ and $c > 3$, we do not explore the general speed of convergence in median polish for multivariate functional data. In our applications, it converges ($r = 2$ and $c > 2$) after two iterations.

2.3 | Median of multivariate functional data

Depending on the choice of depth for multivariate functional data, we might get small differences in the resulting median. All the multivariate functional depths mentioned in Section 1 gave visually unbiased results with small variance. Considering that multivariate functional simplicial depth (MFSD) is equivalent to modified simplicial band depth (MSBD) and the merits of MSBD analyzed by López-Pintado et al. (2014), we adopt modified simplicial band depth (MSBD) to find the median in multivariate functional data. First, we determine the MSBD of each observation in the multivariate functional data; then, we identify the observation corresponding to the largest depth as the median.

Let $D(\cdot; F_Y) : \mathbb{R}^p \rightarrow [0, 1]$ be a stochastic depth function for the probability distribution of a p -variate random vector \mathbf{Y} with cumulative distribution function (cdf) F_Y . Associated with the depth function is the depth region $D_\alpha(F_Y)$ at level $\alpha \geq 0$, defined as $D_\alpha(F_Y) = \{\mathbf{Y} \in \mathbb{R}^p : D(\mathbf{Y}; F_Y) \geq \alpha\}$. For functional data $\mathbf{Y}(t)$, we consider the index $t \in \mathfrak{F}$. Then, the definition of multivariate functional depth (MFD, Claeskens et al., 2014) combines the local depths of $\mathbf{Y}(t)$ at each time point $t \in \mathfrak{F}$ and includes a weight function.

Consider a p -variate stochastic process $\{\mathbf{Y}(t), t \in \mathfrak{F}\}$ on \mathbb{R}^p with cdf $F_{Y(t)}$. Let $C(\mathfrak{F})^p$ be continuous paths for the p -variate process $\mathbf{Y}(t)$, w a weight function that integrates to one on the domain \mathfrak{F} , and D be a general depth function on \mathbb{R}^p . The multivariate functional depth (MFD) of \mathbf{Y} is defined as

$$\text{MFD}(\mathbf{Y}; F_Y) = \int_{\mathfrak{F}} D(\mathbf{Y}(t); F_{Y(t)}) w(t) dt, \quad (3)$$

where the weight function may or may not depend on $F_{Y(t)}$, $t \in \mathfrak{F}$. The weight $w(t)$ can be a constant w independent of time, or it can be proportional to the volume of the depth region, at time point t , which is $w(t) = w_\alpha(t; F_{Y(t)}) = \text{vol}\{D_\alpha(F_{Y(t)})\} / \int_{\mathfrak{F}} \text{vol}\{D_\alpha(F_{Y(u)})\} du$.

Similarly, under a finite sample of multivariate curve observations $\mathbf{Y}(t_j) = \{\mathbf{Y}_1(t_j), \dots, \mathbf{Y}_N(t_j); j = 1, \dots, T\}$ with at each time point t a cdf $F_{Y(t),N}$, the sample multivariate functional depth at $\mathbf{Y} \in C(\mathfrak{F})^p$ is defined with $t_0 = t_1$, $t_{T+1} = t_T$ and $W_j = \int_{(t_{j-1}+t_j)/2}^{(t_j+t_{j+1})/2} w(t) dt$, by

$$\text{MFD}_N(\mathbf{Y}) = \sum_{j=1}^T D(\mathbf{Y}(t_j); F_{Y(t_j),N}) W_j. \quad (4)$$

Following the two definitions of $w(t)$ above, $W_j = w \cdot (t_{j+1} - t_{j-1})/2$ or $W_j = \text{vol}\{D_\alpha(F_{Y(t),N})\}(t_{j+1} - t_{j-1}) / [\sum_{j=1}^T \text{vol}\{D_\alpha(F_{Y(t),N})\}(t_{j+1} - t_{j-1})]$.

The depth function $D(\mathbf{Y}(t); F_{\mathbf{Y}(t)})$ we choose here is the simplicial depth (SD, Liu, 1990) defined as below:

$$\text{SD}(\mathbf{Y}(t); F_{\mathbf{Y}(t)}) = P\{\mathbf{Y}(t) \in \text{simplex}\{\mathbf{Y}_1(t), \dots, \mathbf{Y}_{p+1}(t)\}\}, \quad (5)$$

where $\mathbf{Y}_1(t), \dots, \mathbf{Y}_{p+1}(t)$ represent $p + 1$ independent copies of $\mathbf{Y}(t)$, $\text{simplex}\{\mathbf{Y}_1(t), \dots, \mathbf{Y}_{p+1}(t)\}$ is the simplex with vertices $\mathbf{Y}_1(t), \dots, \mathbf{Y}_{p+1}(t)$ (i.e., to say, $\text{simplex}\{\mathbf{Y}_1(t), \dots, \mathbf{Y}_{p+1}(t)\}$ is the set of all points in \mathbb{R}^p which are convex combinations of $\mathbf{Y}_1(t), \dots, \mathbf{Y}_{p+1}(t)$).

If we use simplicial depth (SD) as the building block in the definition of multivariate functional depth (MFD), then this leads to multivariate functional simplicial depth (modified simplicial band depth, MSBD) in the continuous and finite time domain, respectively:

$$\text{MSBD}(\mathbf{Y}; F_{\mathbf{Y}}) = \int_{\mathfrak{F}} \text{SD}(\mathbf{Y}(t); F_{\mathbf{Y}(t)}) w(t) dt, \quad (6)$$

$$\text{MSBD}_N(\mathbf{Y}) = \sum_{j=1}^T \text{SD}(\mathbf{Y}(t_j); F_{\mathbf{Y}(t_j), N}) W_j. \quad (7)$$

The other four MFD depths can be obtained in a similar way using the corresponding multivariate depths as building blocks in the definition of multivariate functional depth. Here we take the weight to be a constant for convenience of the computation.

2.4 | Hypothesis tests in FMANOVA

Depending on the estimation by median polish or by means, we use either a rank-based test or a mean-based test, to test whether the factor effects are significantly different. We take the two-way functional additive model (2) in Section 2.2 as an example, and tests in the one-way functional additive model (1) in Section 2.1 are simplifications of the example.

The null hypothesis of no row effect is

$$H_{0r} : \boldsymbol{\alpha}_1(t) = \dots = \boldsymbol{\alpha}_r(t) = \mathbf{0}, \text{ vs. } H_{1r} : \text{at least one } \boldsymbol{\alpha}_i(t) \neq \mathbf{0} \quad \forall t \in \mathfrak{F}, i = 1, \dots, r.$$

Similarly, the null hypothesis of no column effect is

$$H_{0c} : \boldsymbol{\beta}_1(t) = \dots = \boldsymbol{\beta}_c(t) = \mathbf{0}, \text{ vs. } H_{1c} : \text{at least one } \boldsymbol{\beta}_j(t) \neq \mathbf{0} \quad \forall t \in \mathfrak{F}, j = 1, \dots, c.$$

2.4.1 | Rank-based test in FMANOVA

We generalize the pairwise nonparametric test in Sun and Genton (2011) to FMANOVA by replacing MBD in the univariate functional data with MSBD in the multivariate functional data:

$$H_0 : \boldsymbol{\mu}_{\mathbf{Y}}(t) = \boldsymbol{\mu}_{\mathbf{Y}'}(t), \quad \forall t \in \mathfrak{F}.$$

Given two multivariate functional samples $\mathbf{Y}_1(t), \dots, \mathbf{Y}_n(t)$ and $\mathbf{Y}'_1(t), \dots, \mathbf{Y}'_m(t)$, denote by $\mathbf{Z}_1(t), \dots, \mathbf{Z}_q(t)$ the new samples selected from one of the two samples (q should be at least $\max(n, m)$, for simplicity, we take $q = \max(n, m)$). The order of $\mathbf{Y}_i(t)$ ($i = 1, \dots, n$) and $\mathbf{Y}'_j(t)$ ($j = 1, \dots, m$) in the new sample can be defined as

$$R(\mathbf{Y}_i) = \frac{1}{q} \sum_{k=1}^q I\{\text{MSBD}_q(\mathbf{Z}_k) \leq \text{MSBD}_n(\mathbf{Y}_i)\}, \quad R(\mathbf{Y}'_j) = \frac{1}{q} \sum_{k=1}^q I\{\text{MSBD}_q(\mathbf{Z}_k) \leq \text{MSBD}_m(\mathbf{Y}'_j)\}.$$

Next, we rank these values from 1 to $n + m$ in an ascending order of R . If we define $W = \sum_{j=1}^m \text{rank}\{R(\mathbf{Y}'_j)\}$, then W under H_0 satisfies the distribution of the sum of m numbers randomly generated from integers 1 to $n + m$ with critical values determined by simulations.

In the one-way model test, $H_0 = H_{0r}$, we can apply the groupwise comparison directly without deducting other factor effects. In the two-way test, to use the above method in testing H_{0r} , we deduct the corresponding column effect from each sample, then divide the samples according to different r row effect levels, and use the groupwise comparisons to test whether at least one α_i is significantly different. In a similar way, to apply the above method in testing H_{0c} , we deduct the corresponding row effect from the original sample, then divide the samples according to different c column effect levels, and use the groupwise comparisons to test whether at least one β_j is significantly different.

2.4.2 | Mean-based tests in FMANOVA

For FMANOVA data, we estimate the functional grand effect and row and column effects by means from $\hat{\mu}_{...}(t)$, $\hat{\mu}_{i..}(t)$, $\hat{\mu}_{j..}(t)$, $\hat{\mu}_{ij..}(t)$, $\hat{\alpha}_i(t)$, and $\hat{\beta}_j(t)$ as below:

$$\begin{aligned}\hat{\mu}_{...}(t) &= \frac{1}{\sum_{i,j} m_{ij}} \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^{m_{ij}} \mathbf{Y}_{i,j,k}(t), \quad \hat{\mu}_{i..}(t) = \frac{1}{\sum_j m_{ij}} \sum_{j=1}^c \sum_{k=1}^{m_{ij}} \mathbf{Y}_{i,j,k}(t), \\ \hat{\mu}_{j..}(t) &= \frac{1}{\sum_i m_{ij}} \sum_{i=1}^r \sum_{k=1}^{m_{ij}} \mathbf{Y}_{i,j,k}(t), \quad \hat{\mu}_{ij..}(t) = \frac{1}{m_{ij}} \sum_{k=1}^{m_{ij}} \mathbf{Y}_{i,j,k}(t) \\ \hat{\alpha}_i(t) &= \hat{\mu}_{i..}(t) - \hat{\mu}_{...}(t), \quad i = 1, \dots, r, \\ \hat{\beta}_j(t) &= \hat{\mu}_{j..}(t) - \hat{\mu}_{...}(t), \quad j = 1, \dots, c.\end{aligned}$$

For each $t \in \mathfrak{F}$, we introduce a total sum of squares matrix

$$\mathbf{T}(t) = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^{m_{ij}} \{\mathbf{Y}_{i,j,k}(t) - \hat{\mu}_{...}(t)\} \{\mathbf{Y}_{i,j,k}(t) - \hat{\mu}_{...}(t)\}^\top.$$

It can be decomposed into three parts:

$$\mathbf{H}_r(t) = \sum_{i=1}^r \sum_{j=1}^c m_{ij} \{\hat{\mu}_{i..}(t) - \hat{\mu}_{...}(t)\} \{\hat{\mu}_{i..}(t) - \hat{\mu}_{...}(t)\}^\top = \sum_{i=1}^r \sum_{j=1}^c m_{ij} \hat{\alpha}_i(t) \hat{\alpha}_i(t)^\top$$

as the row effect sum of squares and cross products matrix,

$$\mathbf{H}_c(t) = \sum_{i=1}^r \sum_{j=1}^c m_{ij} \{\hat{\mu}_{j..}(t) - \hat{\mu}_{...}(t)\} \{\hat{\mu}_{j..}(t) - \hat{\mu}_{...}(t)\}^\top = \sum_{i=1}^r \sum_{j=1}^c m_{ij} \hat{\beta}_j(t) \hat{\beta}_j(t)^\top$$

as the column effect sum of squares and cross products matrix, and

$$\mathbf{E}(t) = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^{m_{ij}} \{\mathbf{Y}_{i,j,k}(t) - \hat{\mu}_{i..}(t) - \hat{\mu}_{j..}(t) + \hat{\mu}_{...}(t)\} \{\mathbf{Y}_{i,j,k}(t) - \hat{\mu}_{i..}(t) - \hat{\mu}_{j..}(t) + \hat{\mu}_{...}(t)\}^\top$$

as the error sum of squares and cross products matrix.

Common test statistics in the MANOVA (Carey, 1998) are as follows, where $l = \{r, c\}$ depends on the significance of the effect tested:

1. Wilks Lambda $\Lambda_l^* = \frac{|\mathbf{E}|}{|\mathbf{H}_l + \mathbf{E}|}$, ($|\cdot|$ denotes the determinant of a matrix)
2. Hotelling–Lawley trace $LH_l = \text{trace}(\mathbf{H}_l \mathbf{E}^{-1})$,
3. Pillai trace $P_l = \text{trace}\{\mathbf{H}_l (\mathbf{H}_l + \mathbf{E})^{-1}\}$,
4. Roy's maximum root $R_l = \lambda_{\max}(\mathbf{H}_l \mathbf{E}^{-1})$.

We reject the null hypothesis for small values of Λ_l^* and large values of LH_l , P_l , and R_l in MANOVA. Extending the sum of squares matrices to FMANOVA (Górecki & Smaga, 2017), we consider their mean behavior among all t values and determine \mathbf{H}_r , \mathbf{H}_c , and \mathbf{E} as follows:

$$\begin{aligned}\mathbf{H}_r &= \int_{\mathfrak{F}} \mathbf{H}_r(t) dt = \sum_{i=1}^r \sum_{j=1}^c m_{i,j} \int_{\mathfrak{F}} \{\hat{\mu}_{i..}(t) - \hat{\mu}_{...}(t)\} \{\hat{\mu}_{i..}(t) - \hat{\mu}_{...}(t)\}^\top dt, \\ \mathbf{H}_c &= \int_{\mathfrak{F}} \mathbf{H}_c(t) dt = \sum_{i=1}^r \sum_{j=1}^c m_{i,j} \int_{\mathfrak{F}} \{\hat{\mu}_{j..}(t) - \hat{\mu}_{...}(t)\} \{\hat{\mu}_{j..}(t) - \hat{\mu}_{...}(t)\}^\top dt, \\ \mathbf{E} &= \int_{\mathfrak{F}} \mathbf{E}(t) dt = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^{m_{i,j}} \int_{\mathfrak{F}} \{\mathbf{Y}_{i,j,k}(t) - \hat{\mu}_{i..}(t) - \hat{\mu}_{j..}(t) + \hat{\mu}_{...}(t)\} \{\mathbf{Y}_{i,j,k}(t) - \hat{\mu}_{i..}(t) - \hat{\mu}_{j..}(t) + \hat{\mu}_{...}(t)\}^\top dt.\end{aligned}$$

In the one-way test, only \mathbf{H} and \mathbf{E} need to be considered. Since it is hard to find even the approximate distribution of the above test statistics under the null hypothesis for FMANOVA, we adopt the permutation tests (Górecki & Smaga, 2017) to determine the critical values. If the test statistics show there exist significant differences among the groups, we may use the parametric tests in Sun and Genton (2012b) to make further inferences by computing the mean in each group and use the functional t -test statistic for pairwise comparison.

3 | SIMULATION STUDY

3.1 | Simulation design

We conduct comprehensive simulation studies to assess the behavior of the proposed functional median polish method using various contaminated models. The comparison with conventional methods illustrates the robustness of our proposal. Specifically, we consider the two-way functional additive model (2) in Section 2.2.

In our case, t is chosen from 50 equally separated points in $[0,1]$. We set the grand effect to be $\boldsymbol{\mu}(t) = (4t, 5t)^\top$, the row effects to be $\boldsymbol{\alpha}_i(t) = ((-1)^i \times 4(t - 0.6)^2, (-1)^i \times 4(t - 0.3)^2)^\top$ for $i = 1, 2$ and the column effects to be $\boldsymbol{\beta}_j(t) = (3(j-2) \cos(2\pi t), 3(j-2) \sin(2\pi t))^\top$ for $j = 1, 2, 3$ and $m_{i,j} = 100$. We set $\epsilon_{i,j}(t) = \mathbf{e}_{i,j}(t)$ when no outliers exist. We choose $\mathbf{e}_{i,j}(t)$ from the bivariate Matérn cross-covariance function (Gneiting, Kleiber, & Schlather, 2010):

$$C_{ii}(s, t) = \sigma_i^2 \mathcal{M}(|s - t|; \nu_i, \eta_i), \quad i = 1, 2,$$

$$C_{ij}(s, t) = C_{ji}(s, t) = \rho_{ij} \sigma_i \sigma_j \mathcal{M}(|s - t|; \nu_{ij}, \eta_{ij}), \quad 1 \leq i \neq j \leq 2,$$

where $\mathcal{M}(h; \nu, \eta) = \frac{2^{1-\nu}}{\Gamma(\nu)} (\eta h)^\nu \mathcal{K}_\nu(\eta h)$, $h = |s - t| \in [0, 1]$, and \mathcal{K}_ν is a modified Bessel function of the second kind of order ν . Here, σ^2 is the marginal variance, $\nu > 0$ adjusts smoothness, and $\eta > 0$ is a scale parameter. For the bivariate Matérn model to be applied, the parameters must be in some valid range. We choose $\sigma_1^2 = \sigma_2^2 = 0.5$, $\rho_{12} = 0.5$, $\nu_{11} = 0.5$, $\nu_{12} = 1.25$, $\nu_{22} = 2$, $\eta_{11} = 2.5$, $\eta_{12} = 2.5$, $\eta_{22} = 2.5$.

Outliers can be generated by transforming $\epsilon_{i,j}(t)$. We consider one clean model, together with two joint and three marginal contamination models.

Model 1 (without contamination)

In Model 1: $\epsilon_{i,j}(t) = \mathbf{e}_{i,j}(t)$, and $\mathbf{e}_{i,j}(t)$ is generated from the Matérn cross-covariance function as stated.

Model 2 (joint magnitude contamination)

Model 2 introduces magnitude outliers for two variables: $\epsilon_{i,j,1}(t) = \mathbf{e}_{i,j,1}(t) + 15c_{i,j}$ and $\epsilon_{i,j,2}(t) = \mathbf{e}_{i,j,2}(t) - 15c_{i,j}$.

We let $c_{i,j}$ be 1 with probability $q_{i,j}$ and 0 with probability $1 - q_{i,j}$. The outlier probability $q_{i,j}$ is different for each cell according to the matrix \mathbf{Q} :

$$\mathbf{Q} = \begin{pmatrix} 0.07 & 0.03 & 0.05 \\ 0.05 & 0.07 & 0.03 \end{pmatrix}.$$

Model 3 (marginal magnitude contamination)

Model 3 only introduces magnitude outliers for Variable 1: $\epsilon_{i,j,1}(t) = e_{i,j,1}(t) + 15c_{i,j}$ and $\epsilon_{i,j,2}(t) = e_{i,j,2}(t)$.

Model 4 (marginal peak contamination)

Model 4 introduces peak outliers for Variable 1: $\epsilon_{i,j,1}(t) = e_{i,j,1}(t) + 15c_{i,j}$ for $t \in [S, S+0.1]$, where S follows a uniform distribution in $[0, 0.9]$ and $\epsilon_{i,j,2}(t) = e_{i,j,2}(t)$.

Model 5 (marginal partial contamination)

Model 5 introduces partial outliers for Variable 1: $\epsilon_{i,j,1}(t) = e_{i,j,1}(t) + 15c_{i,j}$ for $t \in [S, 1]$, where S follows a uniform distribution in $[0.3, 0.9]$ and $\epsilon_{i,j,2}(t) = e_{i,j,2}(t)$.

Model 6 (joint shape contamination)

In Model 6, we introduce shape outliers for both variables. For $r=1$, we introduce outliers by $\epsilon_{i,j,1}(t) = e_{i,j,1}(t) + U\alpha_{i,1} \sin(6\pi t)$; for $r=2$, $\epsilon_{i,j,2}(t) = e_{i,j,2}(t) + U\alpha_{i,2} \cos(6\pi t)$, with outlier probability 0.07 for each variable. When no outlier exists, $\epsilon_{i,j,1}(t) = e_{i,j,1}(t) + U\alpha_{i,1} \sin(2\pi t)$ and $\epsilon_{i,j,2}(t) = e_{i,j,2}(t) + U\alpha_{i,2} \cos(2\pi t)$. Here U follows a uniform distribution on $[-0.3, 0.3]$.

We want to compare the performance of the proposed median polish methods with the classical FMANOVA estimation based on means, where the procedures of median polish include using multivariate functional depth in multivariables and marginal functional depth, considering multivariables independently. The median results are not sensitive to the choice of multivariate functional depth. In this article, we demonstrate the median results based on modified simplicial band depth (MSBD); the remaining results using other multivariate functional depths are provided in the Supplementary Material.

If the estimates using MSBD are better than those using MBD and mean, applying the median polish to multivariate functional data when there is some concern about potential outliers is desirable. We implement 100 independent simulations in each model, and the results are evaluated both marginally and jointly. Marginally, the estimates should provide an accurate description of each component, which is assessed using the functional boxplots as shown in Figures 1 and 2. Jointly, the estimates should preserve the correlation structure among different variables, which are evaluated using the cross-correlation coefficients (Figure 3).

3.2 | Marginal evaluation

Generally, the median polish method makes the results robust but more variable, which can be seen from Model 1 (without contamination). In this model, the estimation obtained from median polish is not as good as the mean estimation, since the latter is unbiased and less variable. For the five contaminated models, we focus on Model 2, in which magnitude outliers for both variables exist. Results for the other models are provided in the Supplementary Material. We choose the grand effect, the first-row effect, and the first-column effect to capture the estimation thanks to the symmetry about 0 in each variable of the model. We get 100 functional medians and 100 mean estimations after 100 simulation replicates, which can be visualized separately through functional boxplots to show their patterns marginally.

Under Model 2 (Figures 1 and 2), there exist significant biases for the grand effect and column effect in the mean estimation. The two median polish methods are not sensitive to the outliers and provide a visually unbiased estimate for each effect. Although the estimation range from MSBD is a bit wider than those from MBD and mean as shown in the functional boxplots, the MSBD and mean-based methods keep the correlation between multivariables, as shown in Section 3.3. Since the median of Variable 1 and the median of Variable 2 chosen from MBD are not from the same multivariate sample, the combined median might deviate from the overall multivariate median, or more severely, be isolated from the whole group of curves. Other contaminated models lead to similar conclusions in the marginal evaluation, except Model 6 (joint shape contamination), where the outliers are not shown in the magnitude. Model 6 does not seem to offer any advantage for the median estimation over the mean estimates from the marginal evaluation, since the mean estimation provides the smallest estimation range among all methods.

3.3 | Joint evaluation

To assess the correlation structure of the estimation from MSBD, MBD, and mean, we calculate the average of the lagged cross-correlation function (ccf) of the two variables among 100 simulations, for each method; according to Algorithm 2,

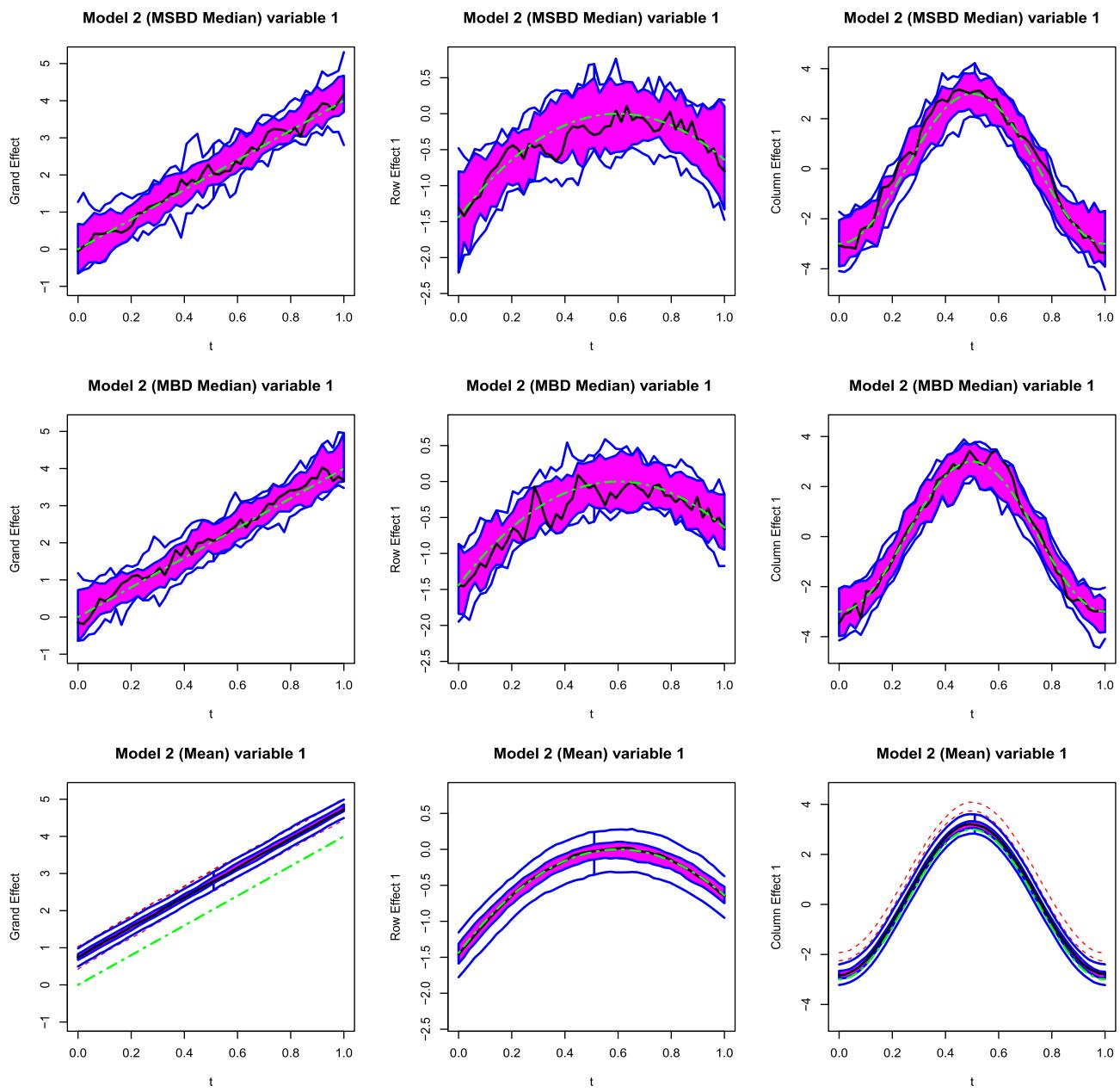


FIGURE 1 Model 2 (joint magnitude contamination), Variable 1: the three rows show functional boxplots of functional median polish using MSBD, functional median polish using MBD, and the classical mean estimation, where the green dashed curve represents the true effect

when row and column effects exist, each effect is a summation of several medians until it does not change. To simplify the iteration times, we remove the grand, row, and column effects, leaving only the errors to infer the average of lagged cross-correlation functions from 100 simulations. Since the errors show no group disparity, we can quickly get the median without the application of Algorithms 1 and 2.

Figure 3 shows the ccf of estimation from MSBD, MBD, and mean, and the black solid curve indicates the theoretical cross-correlation of the random errors. It appears that the estimation from MSBD retains the original correlation structure, the estimate from the mean somewhat less so, whereas the estimate from MBD loses the correlation between variables. Moreover, the combined median obtained from MBD may not come from the original multivariate distribution, and cannot capture the overall median well. Under Model 2, we can conclude that MSBD is a robust method that is stable under outliers and keeps the correlation information, since the mean estimation produces bias and the MBD estimation fails to capture the correlation between multivariabiles.

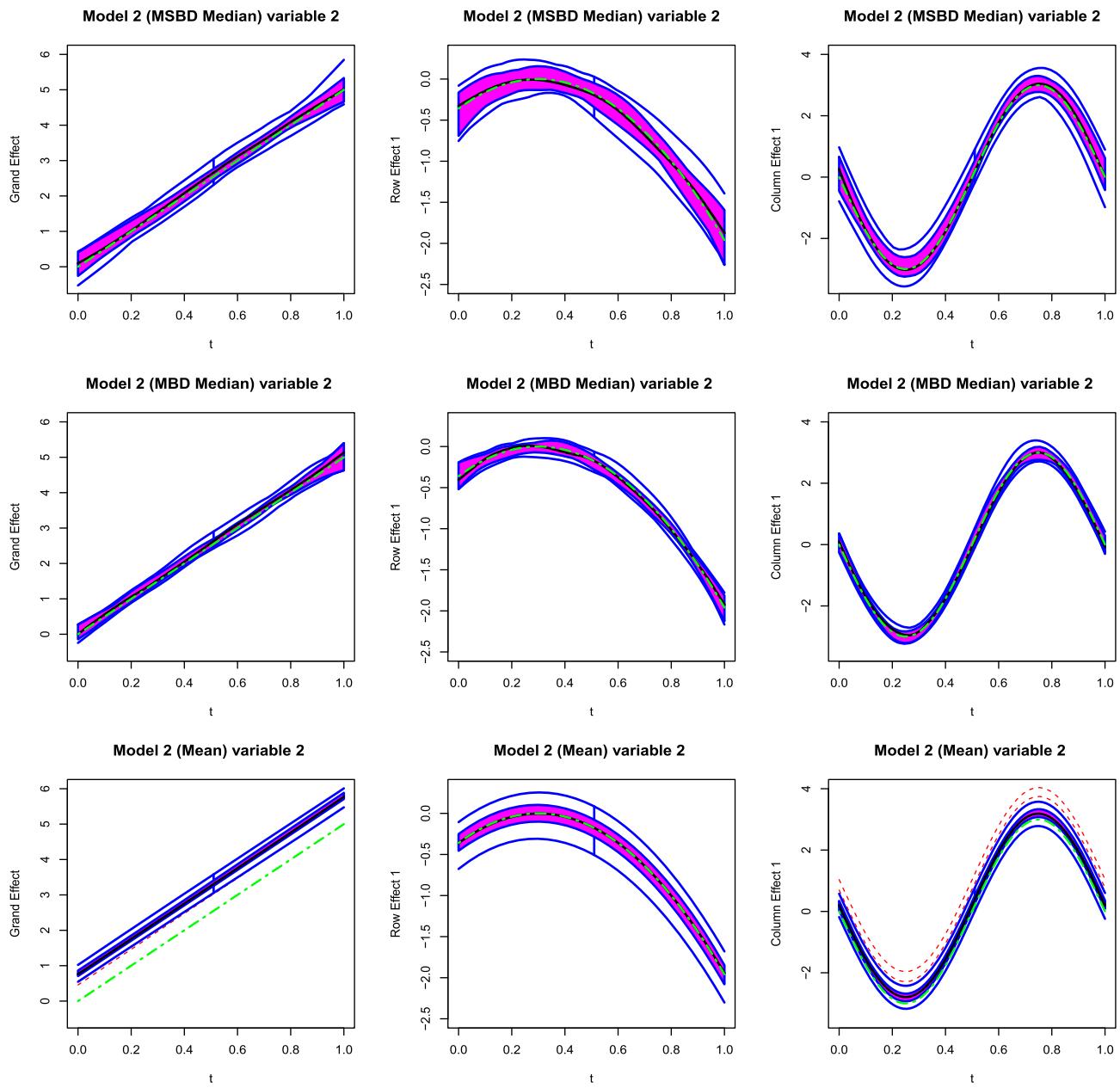


FIGURE 2 Model 2 (joint magnitude contamination), Variable 2: the three rows show functional boxplots of functional median polish using MSBD, functional median polish using MBD, and classical mean estimation, where the green dashed curve represents the true effect

The results from marginal contamination models (Models 3–6) are shown in the Supplementary Material. The cross-correlation structure in the mean estimation is maintained well, except for Models 4 and 5, where the existences of peaks and partial outliers disrupt the structure of cross-correlation in the mean estimation. By contrast, the cross-correlation structure in median estimation based on joint depths is preserved in all the above models.

3.4 | Simulation summary

The comparative drawbacks of the mean estimation with robust FMANOVA are discussed as below. In pure magnitude contaminated models, mean estimation leads to biases with cross-correlation structure maintained; in partial and peak contaminated models, the mean estimation leads to biases and loses some cross-correlation structure. However, in

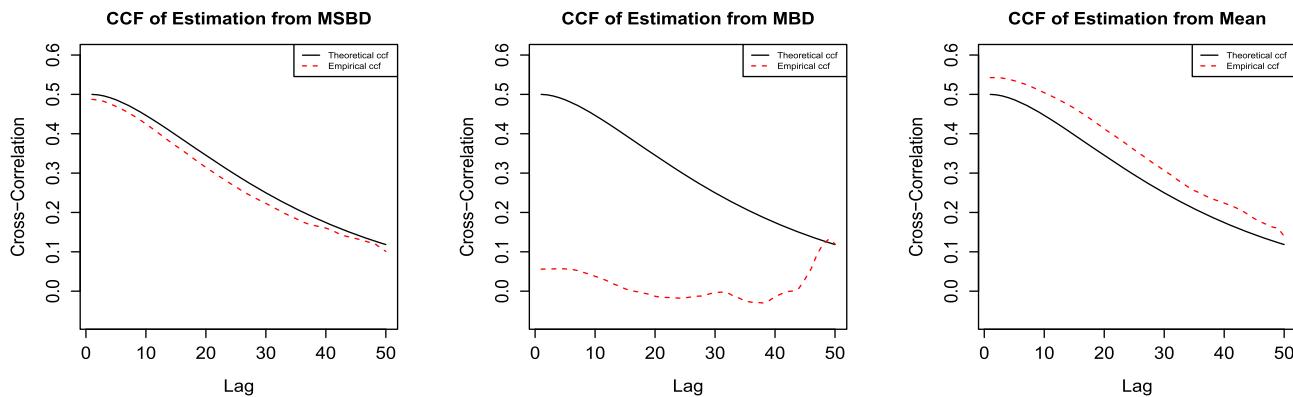


FIGURE 3 The empirical cross-correlation function (ccf) of the estimations from MSBD, MBD and means from Model 2. The red dashed curve shows the empirical ccf and the black solid curve shows the theoretical ccf

shape-contaminated cases, the mean estimation is better than the robust FANOVA with the least variable and visually unbiased estimation range as well as maintaining cross-correlation structure. In the above cases, the cross-correlation structure using the MBD method is lost, whereas it remains in the robust FANOVA. Combining both marginal and joint evaluations, the robust FANOVA leads to better results in most cases, except for the case with marginal contamination.

So far, we illustrated the robustness of FANOVA when the observations are independent and smooth. However, in the application of robust FANOVA to spatial statistics problems, it is difficult to require that data be entirely independent and smooth, as observations from nearby locations may share some features, and some raw climate data, such as precipitation, are usually not smooth. To show that the robustness is not significantly affected in moderately dependent data and nonsmooth data cases, we discuss the effects on median polish of nonindependent but smooth data, and independent but nonsmooth data.

In Example 1, we introduce a slight dependence in the errors according to the functional autoregressive model of order one (FAR(1), Martínez-Hernández, Genton, & González-Farías, 2019) while smoothness is maintained. For any fixed i, j , $n = 1, \dots, m_{i,j}$, we add an observation index n to express the new error $\epsilon_{i,j}^{(n)}(t)$ due to the correlation between observations:

$$\epsilon_{i,j}^{(n)}(t) = \int_{\mathfrak{F}} \beta(s, t) \epsilon_{i,j}^{(n-1)}(s) ds + \epsilon_n(t), \quad n = 1, \dots, m_{i,j},$$

where $\beta(s, t) = \exp\{-(s^2 + t^2)/2\}$ represents the kernel with $\{\int_{\mathfrak{F}} \int_{\mathfrak{F}} \beta^2(s, t) ds dt\}^{1/2} < 1$, and $\epsilon_n(t)$ has the Matérn cross-covariance function. Then, we repeat Model 1 for the nonindependent data set, and find that the results are similar to those with independent data, except that the estimation range with nonindependent data is wider. This is because when observation data are not independent, the effective sample size is smaller; hence, the uncertainty is more significant. Likewise, in Example 2, by only adjusting smoothness parameters in the Matérn cross-covariance model (set $v_1 = 0.05$, $v_2 = 0.1$, and $v_{12} = 0.075$), we consider the nonsmooth but independent data case in Model 1 to see whether the robust FANOVA only applies to smooth data. It turns out that although the estimation ranges of nonsmooth data are larger than estimation ranges of smooth data from marginal evaluation, the empirical ccf with nonsmooth data is close to the theoretical ccf in the joint evaluation. In Examples 1 and 2, we observe that the robustness of our method is not seriously affected in the moderate dependence and nonsmooth scenarios, at the cost of wider estimation ranges in the marginal evaluation and rougher cross-correlation functions in the joint evaluation.

Overall, in both joint and marginal contamination cases considered, as long as a moderate correlation between two variables exists, using MSBD leads to robust and visually unbiased results maintaining the correlation structure. Moreover, when shape contaminations exist, the advantage of robust FANOVA with median estimates over mean estimation is not apparent. Based on the discussion above and the results from the other models, we recommend median polish using MSBD as an alternative method to the mean-based FANOVA when the practitioner has concerns about abnormal observations.

TABLE 1 *p*-values of mean-based and rank-based tests for the Saudi Arabia wind speed data

<i>p</i>-values	Mean-based test
First stage	.029
Second stage: <i>U</i>	.043
Second stage: <i>V</i>	.015
<i>p</i>-value	Rank-based test
Bivariate	.037

4 | ENVIRONMENTAL APPLICATIONS

Functional additive models can be applied to environmental data to analyze disparities of dependent variables in various levels of factors. In this section, we use three real data sets from environmental science and compare the representation based on the functional median polish with mean-based counterparts. The Saudi Arabian wind image data set in Section 4.1 is an instance of a one-way bivariate FANOVA visualized by heat maps, and the Canadian and Spanish spatial weather data in the next two subsections are examples of one-way bivariate FANOVA and two-way trivariate FANOVA, respectively, shown with functional boxplots.

4.1 | Saudi Arabia bivariate wind speed data

We apply the functional median polish to daily wind-speed data from NOAA (National Oceanic and Atmospheric Administration), using the GFDL-ESM2M model based on historical wind speeds during 1976–2005. This model was used by Chen, Castruccio, Genton, and Crippa (2018) to analyze the wind-power potential in Saudi Arabia. The project is based on the Coordinated Regional Climate Downscaling Experiment (CORDEX) MENA-22 (Middle East North Africa) domain and the grid resolution is 0.22 by 0.22 degrees with a rotated pole system where one quasi-uniform resolution represents approximately 25 km. The response variables are two wind-speed components: *U* (zonal velocity) and *V* (meridional velocity) at 10 m above the displacement height. Considering that averaging the data over different years directly loses much information because of the existence of phase and amplitude variation, we implement the time warping by Srivastava, Wu, Kurtek, Klassen, and Marron (2011) and Tucker, Wu, and Srivastava (2013) for the data in each location, respectively. Then, for each day, we average the data over these 30 years and get 365 daily wind speed images. Given the climate in Saudi Arabia, we divide the wind data into two parts: speed in the dry season (April to October) and speed in the wet season (November to March).

After implementing the one-way bivariate FANOVA algorithm in Section 2.1, we obtain Figure 4, which presents the seasonal effects. Generally, the median and mean estimations in both *U* (eastward wind) and *V* (northward wind) components show some common patterns, except that the mean estimation tends to be smoother since the mean considers all observations and averages them. Observations of positive and negative values exist when one area experiences winds with reverse directions for various times of the year. Hence, median estimates of different components of wind speed provide a more intuitive overview of the wind direction and the wind speed across all of Saudi Arabia. According to Table 1, both the rank-based test and the mean-based test reject H_0 at the 5% significance level; this means that there exist significant disparities between the wind speed in wet and dry seasons. As we cannot get a clear view of the wind speed after decomposing the data into the grand and season effects, we obtain the fitted wind speed by applying model (1) in Section 2.1 and combining the grand effect and the seasonal effect together. We compare the fitted median polish estimation and the fitted mean estimation in Figure 5. The pattern of the median estimation resembles that estimated by means except in some boundary regions where the wind speed in the median polish version is more significant. Typically, in the dry season, for the *U* component, large values appear more evident in a wider range of areas in Western Saudi Arabia, which shows that median polish provides a more precise method to estimate Saudi Arabia's wind speed.

By combining the two components, we realize that the prevailing wind direction is northeast in the dry season and southeast in the wet season in Saudi Arabia. The prevailing wind direction in different seasons partially explains why the climate in Saudi Arabia is semiarid to arid even though it is mostly surrounded by the sea. The northeast trade winds in the dry season come from arid land masses, such as the Iranian Plateau. When the offshore wind blows from the continent

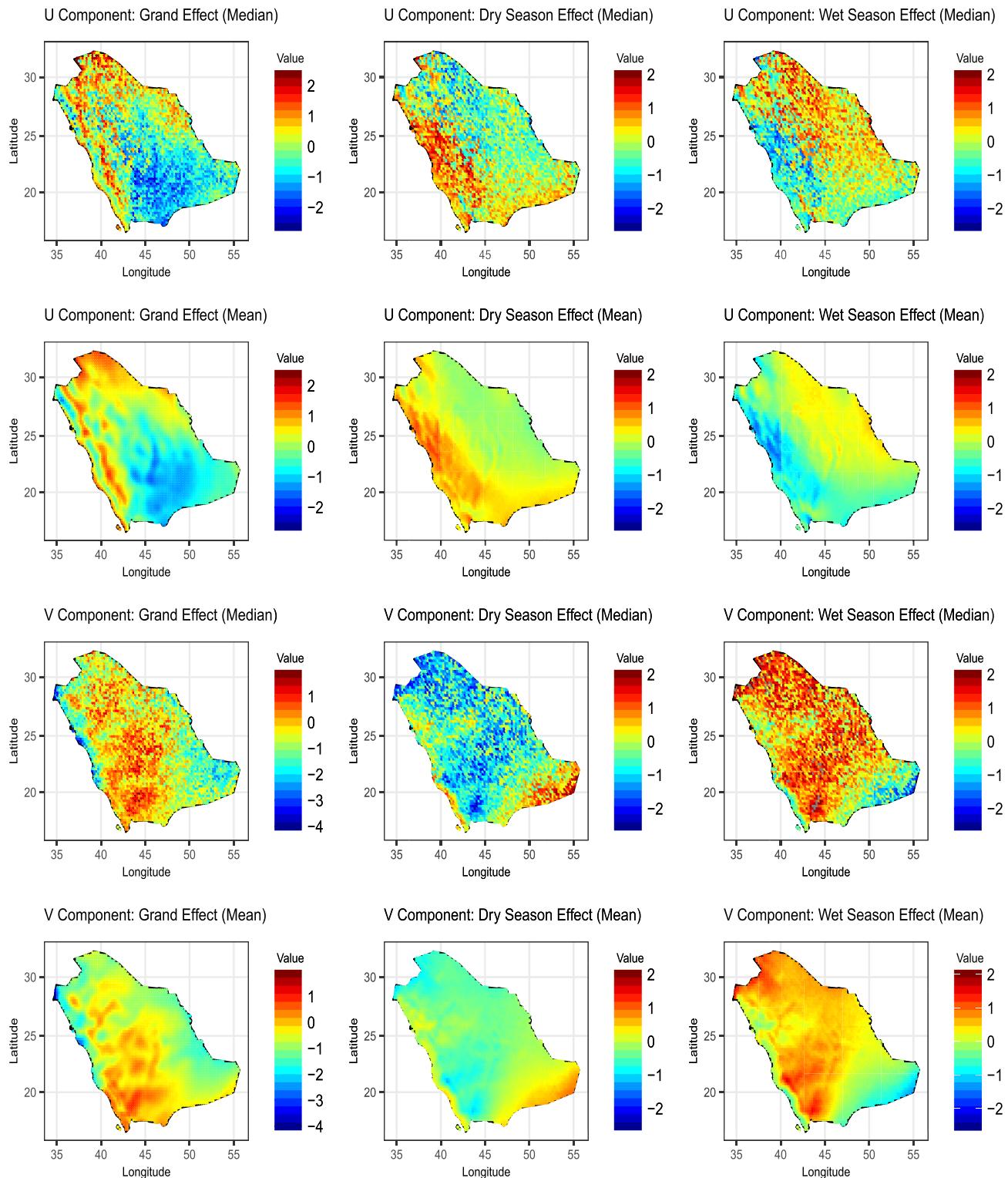


FIGURE 4 Saudi Arabian wind speed: comparison of estimation of the functional grand and seasonal effects using both functional median polish and using mean FMANOVA. A positive U component is eastward wind, and a positive V component is northward wind. The dry season is from April to October, and the wet season is from November to March

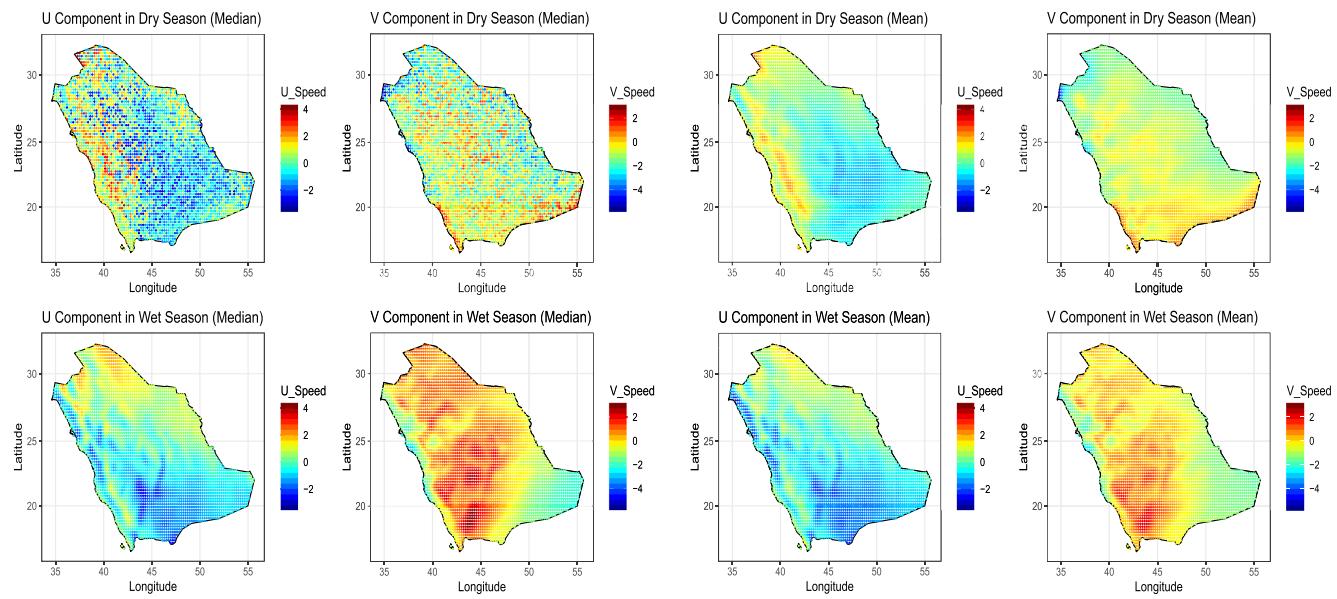


FIGURE 5 Saudi Arabian fitted wind speed (adding the functional grand effect and functional seasonal effects). The four plots on the left are median estimations, and the four plots on the right are mean estimations. A positive U component is eastward wind and a positive V component is northward wind. The dry season is from April to October and the wet season is from November to March

to the sea, on one hand, the wind does not bring much water vapor, but on the other hand, the water vapor does not condense into rain and moves to the ground at higher temperatures under the influence of the subtropical high pressure. On the contrary, the southeast wind in the wet season is mainly from the Ethiopian Highlands and the Indian Ocean. If the wind from the sea is stronger than that from arid Ethiopian Highlands, it may carry much water vapor and turning the vapor into rain becomes possible. Still, the yearly average precipitation in Saudi Arabia is less than 200 mm and the wet season every year is quite short; even if the wind from the Indian Ocean arrives in Southern Saudi Arabia, factors such as air pressure, ocean currents, topography, vegetation coverage, and human activity affect the formation of rain. In general, wind from the south is moister compared with that from the north in Saudi Arabia.

It is worth noting that not all locations in Saudi Arabia share the same wind direction. For example, when most of the areas are under the influence of the northeast wind in the dry season, the southern boundary area is influenced by the southeast wind from the mountainous regions on the north side of the Gulf of Aden while the west part of Saudi Arabia is under the influence of the west wind from the Red Sea. The terrain of Saudi Arabia is high in the west and low in the East, with a long narrow plain in the far western area close to the Red Sea. Although there is wind from the Red Sea toward the mainland, it is limited to the plain in the west and mostly cut off by the Arabian Highlands, which prevent the wind from going deep into the middle of Saudi Arabia.

4.2 | Canadian bivariate weather stations data

We consider the weather data from Canada analyzed by Ramsay and Silverman (2005) and Sun and Genton (2012b). We choose both temperature and log-precipitation data from 35 weather stations located in four climatic regions: the Arctic, Atlantic, Continental, and Pacific averaged over 1960 to 1994. The data are presented in Figure 6, together with their partitions. According to the one-way functional additive model (1) in Section 2.1, we perform median polish as well as the classical mean version to estimate the grand and regional effects illustrated in Figure 7.

The functional grand effect shows that both the temperature and the log-precipitation tend to fluctuate, and that the functional regional effects consist of different temperature and log-precipitation patterns of the four climatic regions. As shown in the left column of Figure 7, the overall tendency of temperature to fluctuate in the mean version of FMANOVA is smoother and less fluctuating in Winter and Spring. In addition, the whole trend of the log-precipitation pattern in the mean version is flatter and higher. Also, we note that the regional effects for the Continental area are 0 for both the

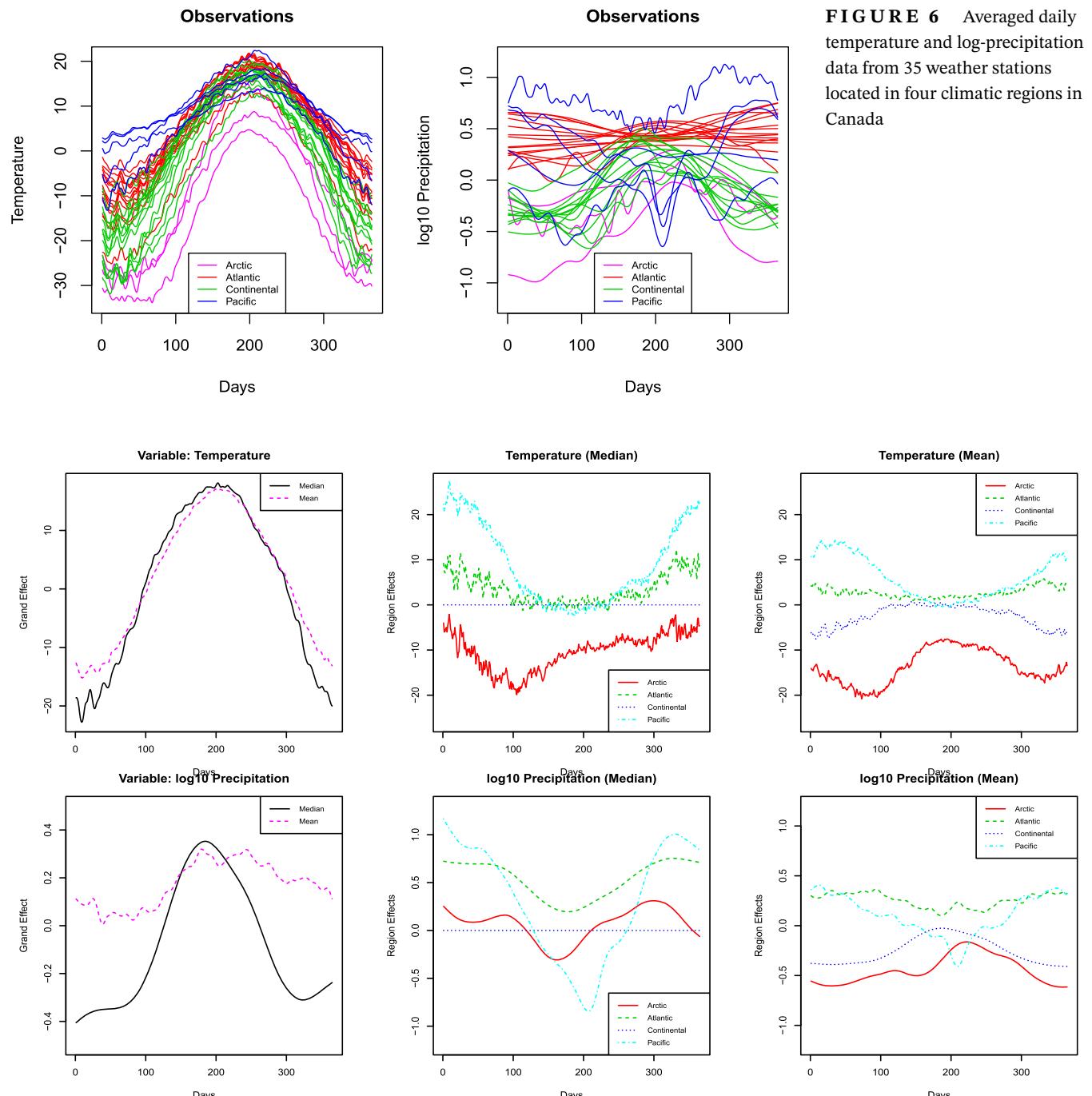


FIGURE 7 Canadian weather: comparison of estimations of the functional grand effect and functional region effects using functional median polish and using mean FMANOVA

temperature and the log-precipitation, which means that the grand effect describes an overall tendency in the Continental Area. When we compare the divergence of the grand effect in the median polish and the mean versions, we compare potential abnormal observations in other areas than those in the Continental area. The left plot in Figure 6 shows that, in the Pacific and Atlantic, the temperatures during the Winter and Spring months are higher than the temperatures in the Continental region. The right plot in Figure 6 shows that all observations in the Atlantic region and one observation in the Pacific region in log-precipitation are quite flat between 0.2 and 1, whereas observations in the Continental area fluctuate between -0.5 and 0.5, which makes the grand effect in the mean version flatter and above most of the grand effect from the functional median polish.

TABLE 2 *p*-values of mean-based and rank-based tests for the Canadian data

<i>p</i>-values	Mean-based test					
First stage	<i>p</i> -value					
Second stage	Ar-At	Ar-C	Ar-P	At-C	At-P	C-P
Temp	.032	.043	.034	.043	.047	.036
Precip	.024	.144	.041	.045	.090	.048
<i>p</i>-values	Rank-based test					
Bivariate	.015	.036	.022	.041	.038	.047

Note: Ar, At, C, and P represent Arctic, Atlantic, Continental, and Pacific areas, respectively. Values in bold mean the group differences are not significantly different at 5% level.

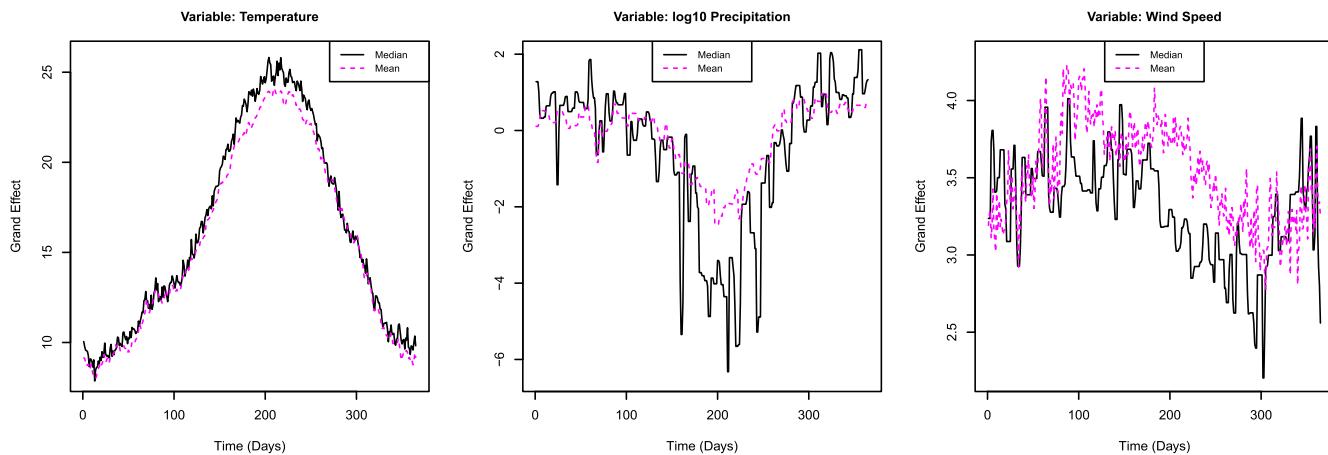


FIGURE 8 Spanish weather: comparison of estimation of the functional grand effects using functional median polish and using mean FMANOVA

Then, we show the regional effects of temperature and log-precipitation according to the median estimation (middle column in Figure 7). The Continental region remains the same in the functional grand effect of temperature and log-precipitation, using the median polish estimation. Also, the Atlantic area tends to be warmer and has more precipitations overall. The Pacific region is warmer with heavier rainfalls, especially during the Winter season. The Arctic region is always colder particularly in the Winter and Spring, and the level of precipitations is a bit higher than that in the Continental area and more substantial in most seasons of the year except for Summer. As a comparison, regional effects fitted by means (right column in Figure 7) fluctuate less and closer to each other, especially during Winter and Spring for both temperature and log-precipitation. Therefore, we tend to get bigger *p*-values that do not show significant differences of regional effects in the mean version.

From the first stage mean-based tests in Table 2, we reject H_0 at the 5% significance level in one-way FMANOVA, suggesting that there exists at least one regional difference in temperature or log-precipitation. The second stage mean-based tests in Table 2 show that in precipitation, regional effects between Arctic and Continental and between Atlantic and Pacific are not significantly different at the 5% significance level. However, median-based tests show all the regional effects are significantly different at the 5% significance level. Therefore, functional median polish provides a more robust estimation of the factor effects free from outliers.

4.3 | Spanish trivariate weather data

We analyze the Spanish weather data studied by Dai and Genton (2018). We choose the daily temperature, log-precipitation, and wind speed data averaged over 1980–2009 from 73 weather stations and consider the effects of location and altitude with four levels and two levels, respectively. We perform a two-way functional median polish as well as

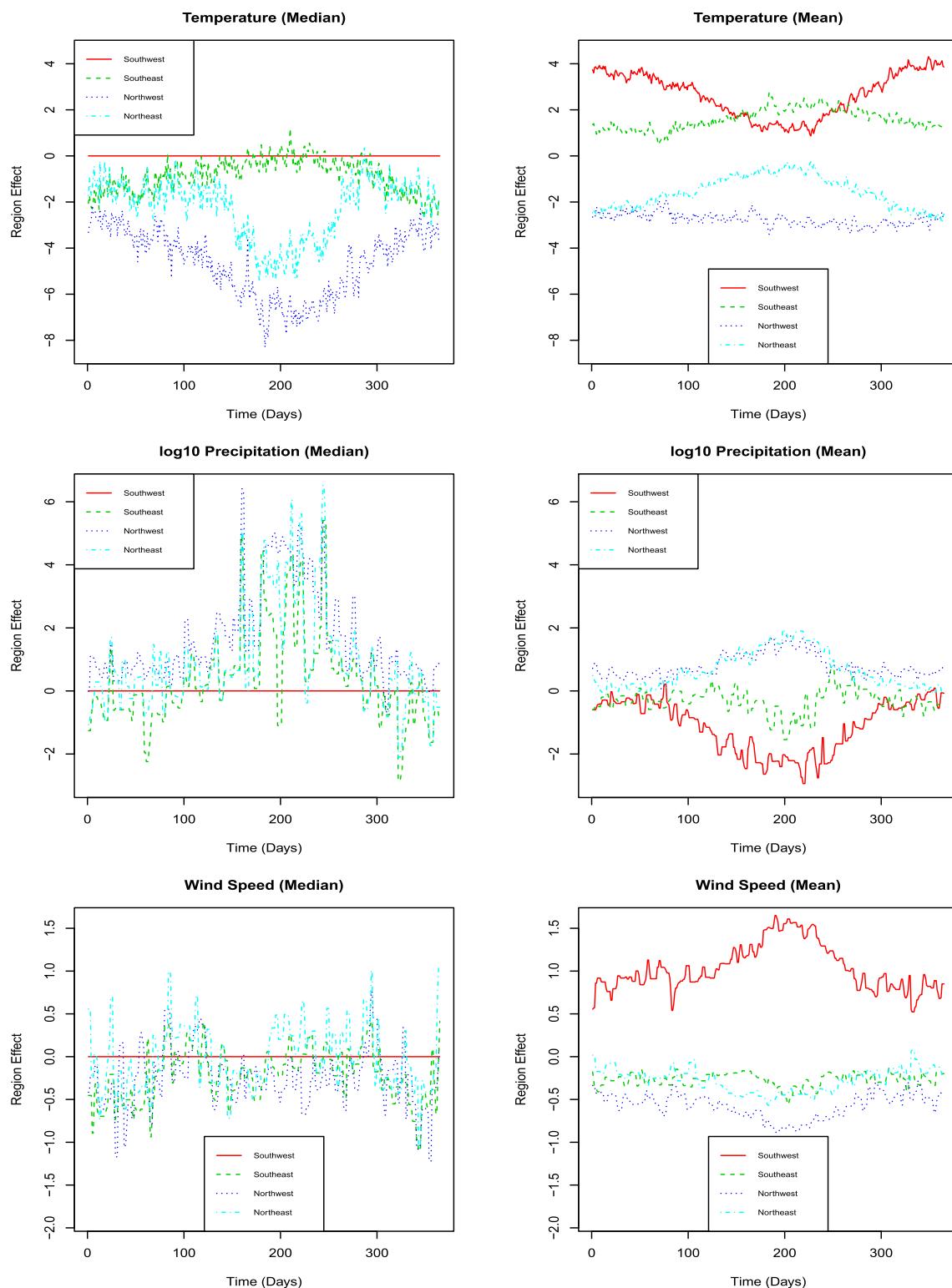


FIGURE 9 Spanish weather: comparison of estimation of the functional regional effects using functional median polish and using mean FMANOVA

TABLE 3 *p*-values of mean-based and rank-based test for the Spanish weather data

<i>p</i>-values	Mean-based test					
First stage	.031					
Second stage: altitude	Low-high					
Temp	.018					
Log precip	.036					
Wind speed	.047					
Second stage: region	SW-SE	SW-NE	SW-NW	SE-NE	SE-NW	NE-NW
Temp	.065	.019	.023	.017	.044	.040
Log precip	.042	.035	.031	.032	.037	.263
Wind speed	.014	.002	.008	.187	.644	.330
<i>p</i>-values	Median-based test					
Altitude	Low-high					
Trivariate	.022					
Region	SW-SE	SW-NE	SW-NW	SE-NE	SE-NW	NE-NW
Trivariate	.070	.035	.041	.142	.032	.469

Note: SW, SE, NE, and NW represent southwest, southeast, northeast, and northwest areas, respectively. Values in bold mean the group differences are not significantly different at 5% level.

the mean FMANOVA version to estimate the grand, the regional and the altitude effects shown in Figures 8, 9, and 11 for model (2) in Section 2.2.

From Figure 8, we can see that the grand effect of temperature for the median polish and for the mean version are quite similar, whereas the grand effects of log-precipitation and wind speed in median polish fluctuate more than those in the mean version.

Mean-based tests from the first stage in Table 3 reject H_0 in two-way FMANOVA at the 5% significance level from the first stage, which means that at least one variable shows a difference in some factor levels. Then, we do pairwise comparisons to find if some group behaves significantly differently. We start with regional effects. As shown in the first row of Figure 9, temperature in the southwest region is zero on the top representing the overall tendency, followed by the southeast, northeast, and northwest in the median polish, whereas the four areas have disparities in the mean estimate. For log-precipitations, regional effects in the median polish estimation show that the northwest region has higher pattern, followed by the northeast and southeast; the southwest region is 0 and represents the overall tendency. However, for log-precipitations, regional effects in the mean version are smoother and fluctuate less. As shown in the last row of Figure 9, for the wind speed variable, regional effects from the median polish do not show any visible difference, but regional effects from the mean estimation show that the southwest regional effect is significantly larger than the other three regional effects. Hence, in Table 3, for wind speed, mean-based tests suggest that the southwest region is significantly different from the other areas at the 5% significance level.

As shown in Figure 10, the wind speed samples in the southeast, northwest, and northeast are almost the same. In the southwest, the curves on the bottom have similar patterns as those in other areas, but the six curves on the top show more massive fluctuations. It is likely that the median is still close to the median of other areas, so median regional effects in the wind speed do not show much difference. In the mean estimation, the regional effect in the southwest is influenced by the top curve observations and shows a higher value than that in other areas.

Figure 11 shows the altitude effects of the temperature, log-precipitation, and wind speed in both median and mean versions. Although we reject H_0 when we make a pairwise comparison of altitude effects of log-precipitation in mean-based tests from Table 3, or a comparison of the effects of altitude in trivariate median-based tests, it is worth noting that the log-precipitation has different trends in the mean and median polish versions. In the mean version, the low-altitude effect is smaller than the high-altitude effect during Summer and Autumn, but the median polish shows that the low-altitude effect is smaller and keeps swinging during the whole year.

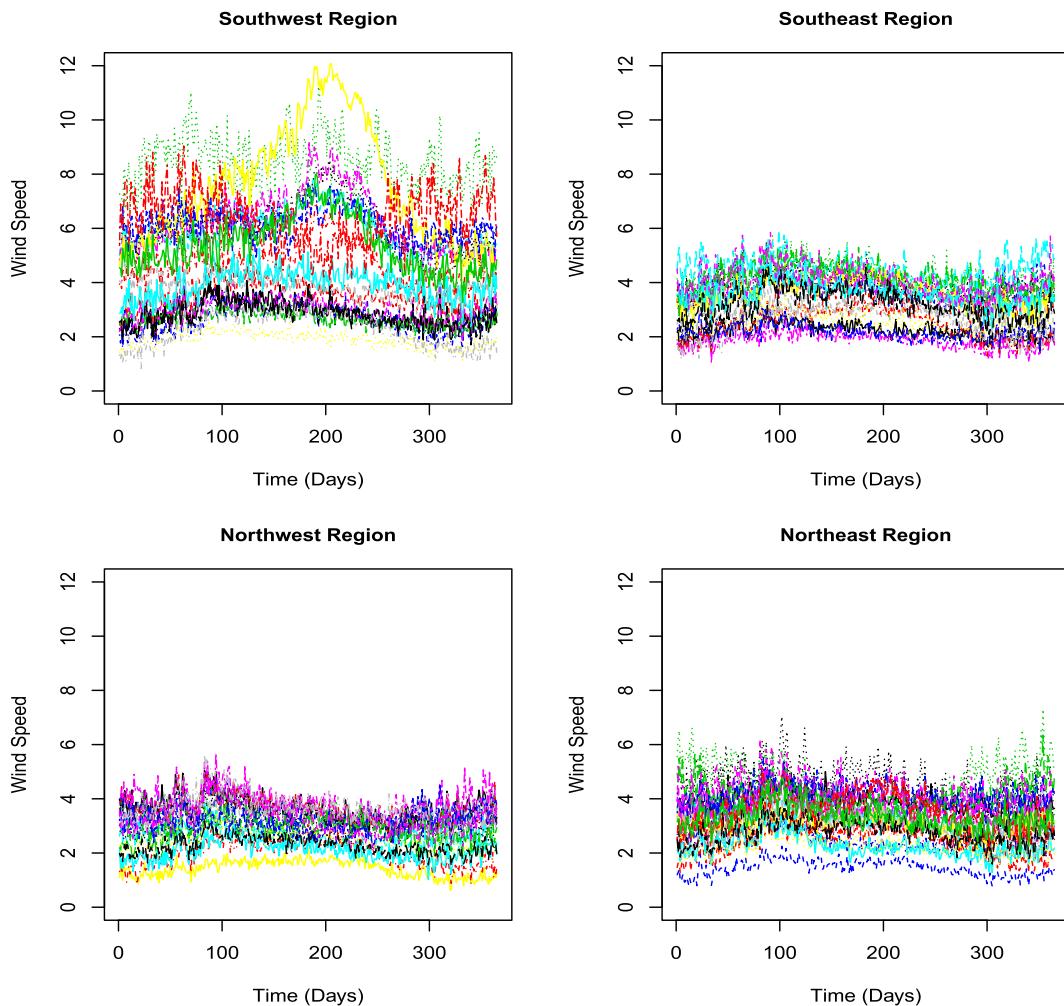


FIGURE 10 Spanish weather: averaged daily wind speed data at 73 weather stations from four climate regions

Figure 12 shows that the observations have different fluctuation tendencies between low and high altitude regions, where some fluctuate a lot and others fluctuate a little. In low-altitude areas, there are more observation samples, where two curves are quite steep and flat for the whole year but more observations swing and stay low in Summer and Autumn. In high-altitude regions, two curves are above the majority and are quite flat, but most curves keep fluctuating during the whole year. Since the mean is not resistant to outliers, the altitude effects estimated by traditional means combine both the flat and fluctuating shapes, which tend to be smoother than effects from median polish. In addition, the low-altitude effect is shifted downward by abnormal observations on the bottom, and the high altitude effect is moved upward by outliers on the top in the mean version.

5 | DISCUSSION

This article, combining the advantages of median polish with the use of multivariate functional depth, considered multivariate functional data and proposed functional median polish in FMANOVA in order to determine the median estimates. The iterative algorithm and the hypothesis tests were respectively generalized to perform the estimation and to check whether the factor effects were significant. The functional median polish, compared with the classical mean estimate, is a more robust method to fit the additive one-way and two-way FMANOVA models because it is less influenced by outliers.

As an extension of the functional median polish from Sun and Genton (2012b), we generalized the algorithm for one-way and two-way FMANOVAs and the hypothesis tests where no interaction term appears. Our simulation studies showed that the functional median polish was robust under the existence of magnitude and shape outliers, whereas the classical mean version of FMANOVA was not resistant to outliers. Magnitude (pure, peak, and partial contamination)

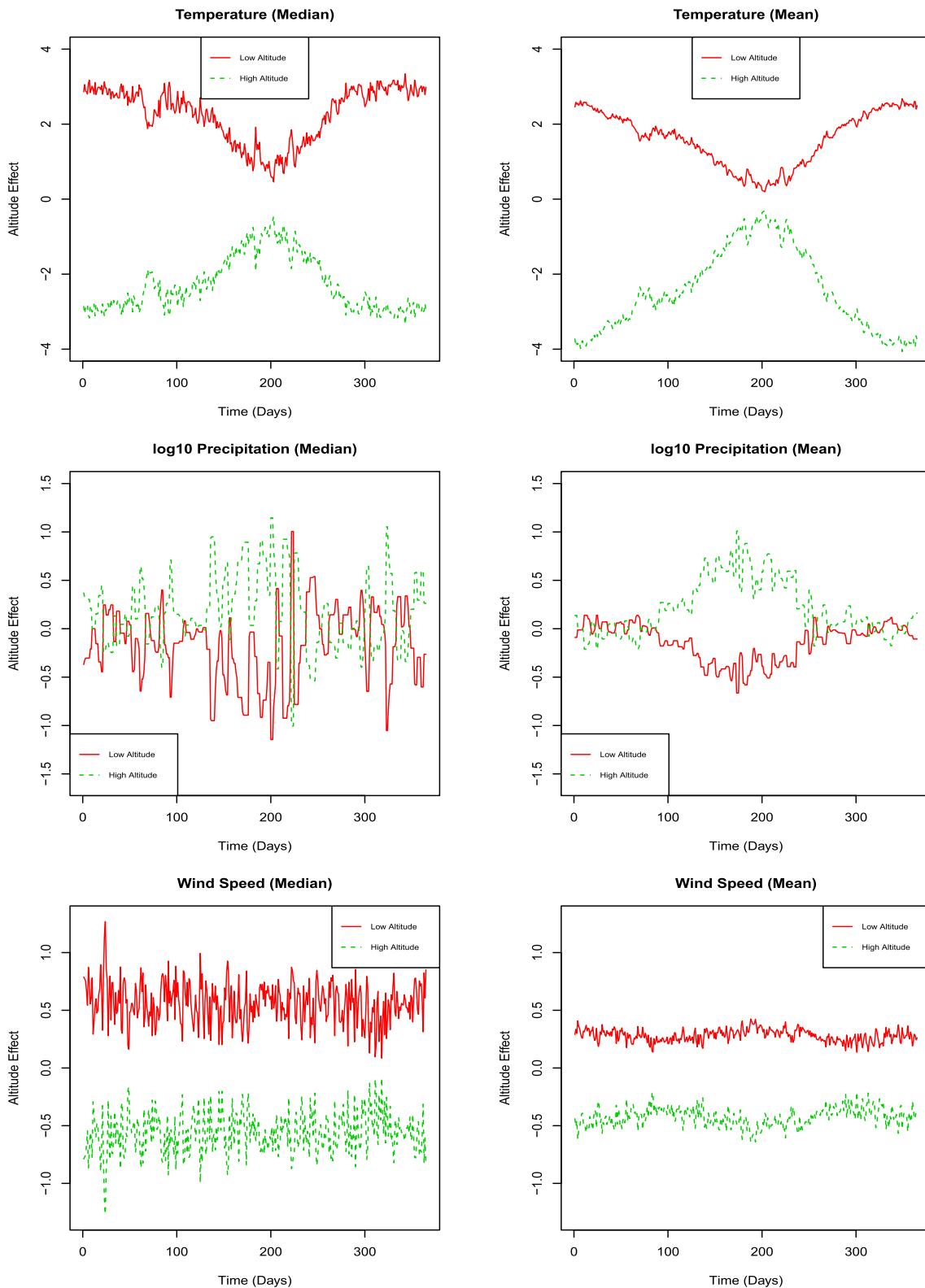


FIGURE 11 Spanish weather: comparison of estimation of the functional altitude effects using functional median polish and using mean FMANOVA

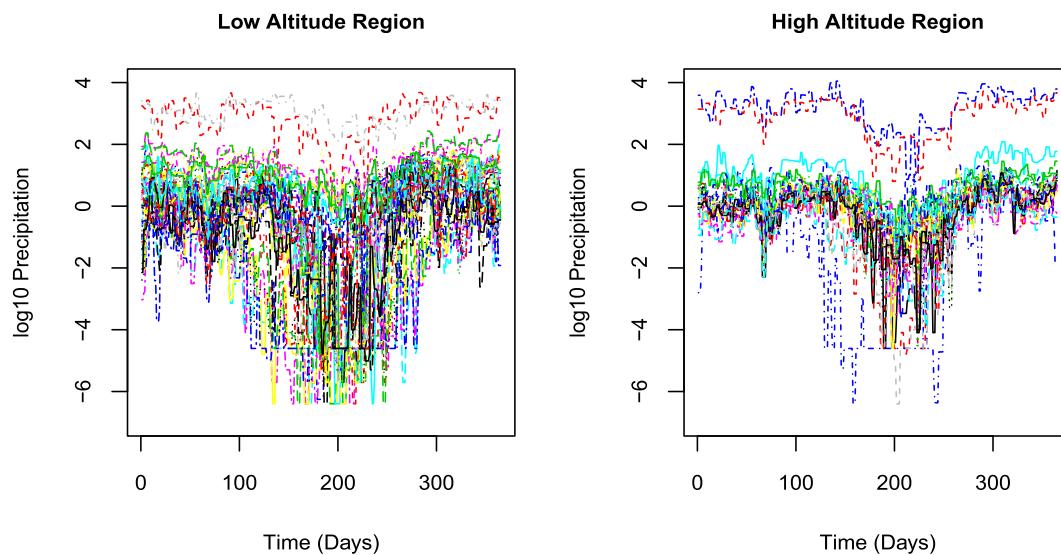


FIGURE 12 Spanish weather: averaged daily log-precipitation data at 73 weather stations from four climate regions

outliers produced some biases in the mean estimation. Also, the cross-correlation structure of the mean estimation was not maintained in the peak and partial contamination. However, we found that the shape contamination neither brought apparent biases nor disrupted the cross-correlation structure for the mean estimation. Hence, the median polish in the magnitude case had an advantage over the mean estimate, but the benefit of median polish was not shown in the shape contamination case. Furthermore, by using nonindependent and nonsmooth data in the FANOVA model separately, we noticed in our examples that neither independence nor smoothness was necessary for the robust FANOVA to be applied.

We found that the determination of the median in multivariate functional data was not sensitive to the choice of multivariate functional depth, which displayed the generalizability of the median polish in FANOVA. We extended the hypothesis rank-based test for the median polish to multivariate functional data. Therefore, in the multivariate functional additive model framework, alongside the estimation of the grand effect and factor effects by functional median polish, the response variables could be fitted by summation of the grand effect and factor effects. Hypothesis tests could be constructed where nonparametric rank-based tests are compared with mean-based tests.

The applications consisted of three environmental data collected from Saudi Arabia, Canada, and Spain, respectively. The first application (Saudi Arabia) involved functional spatial data, and the latter two were functional temporal cases. From all examples, the functional median polish provided robust estimates of the grand effect and factor effects as an alternative to FANOVA fitted by means.

ACKNOWLEDGEMENT

This research was supported by the King Abdullah University of Science and Technology (KAUST) and National Natural Science Foundation of China (NSFC, No.11901573).

ORCID

Zhuo Qu  <https://orcid.org/0000-0001-5805-6101>

REFERENCES

- Berke, O. (2001). Modified median polish kriging and its application to the Wolfcamp–Aquifer data. *Environmetrics*, 12(8), 731–748.
- Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance*, Quantitative applications in the social sciences, (Vol. 54). Newbury Park, CA: Beverly Hills, Calif. : Sage Publications. <https://library.isical.ac.in/cgi-bin/koha/opac-detail.pl?biblionumber=301459>.
- Carey, G. (1998). Multivariate analysis of variance (MANOVA): I theory. Academic Press, Boston.
- Chen, S., & Farnsworth, D. (1990). Median polish and a modified procedure. *Statistics & Probability Letters*, 9(1), 51–57.
- Chen, W., Castruccio, S., Genton, M. G., & Crippa, P. (2018). Current and future estimates of wind energy potential over Saudi Arabia. *Journal of Geophysical Research: Atmospheres*, 123(12), 6443–6459.
- Claeskens, G., Hubert, M., Slaets, L., & Vakili, K. (2014). Multivariate functional halfspace depth. *Journal of the American Statistical Association*, 109(505), 411–423.

- Cressie, N. A. (1984). Median polish kriging. *Biometrics*, 40(4), 1187.
- Cuesta-Albertos, J. A., & Nieto-Reyes, A. (2008). The random Tukey depth. *Computational Statistics & Data Analysis*, 52(11), 4979–4988.
- Cuevas, A., Febrero, M., & Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics*, 22(3), 481–496.
- Dai, W., & Genton, M. G. (2018). Multivariate functional data visualization and outlier detection. *Journal of Computational and Graphical Statistics*, 27, 923–934.
- Dai, W., & Genton, M. G. (2019). Directional outlyingness for multivariate functional data. *Computational Statistics & Data Analysis*, 131, 50–65.
- Emerson, J. D., & Hoaglin, D. C. (1983). Analysis of two-way tables by medians. *Understanding Robust and Exploratory Data Analysis*, 165–210.
- Fink, A. (1988). How to polish off median polish. *SIAM Journal on Scientific and Statistical Computing*, 9(5), 932–940.
- Fraiman, R., & Muniz, G. (2001). Trimmed means for functional data. *TEST*, 10(2), 419–440.
- Genton, M. G., Johnson, C., Potter, K., Stenchikov, G., & Sun, Y. (2014). Surface boxplots. *Stat*, 3(1), 1–11.
- Gneiting, T., Kleiber, W., & Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105(491), 1167–1177.
- Górecki, T., & Smaga, Ł. (2017). Multivariate analysis of variance for functional data. *Journal of Applied Statistics*, 44(12), 2172–2189.
- Hubert, M., & Van der Veeken, S. (2008). Outlier detection for skewed data. *Journal of Chemometrics*, 22(3-4), 235–246.
- Ieva, F., & Paganoni, A. M. (2013). Depth measures for multivariate functional data. *Communications in Statistics-Theory and Methods*, 42(7), 1265–1276.
- Kaufman, C. G., & Sain, S. R. (2010). Bayesian functional ANOVA modeling using Gaussian process prior distributions. *Bayesian Analysis*, 5(1), 123–149.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). Two-factor studies with equal sample size. *Applied linear statistical models* (Vol. 5). 5th ed. 817–827. Boston, MA: McGraw-Hill Irwin.
- Liu, R. Y. (1990). On a notion of data depth based on random simplices. *The Annals of Statistics*, 18, 405–414.
- López-Pintado, S., & Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association*, 104(486), 718–734.
- López-Pintado, S., Sun, Y., Lin, J. K., & Genton, M. G. (2014). Simplicial band depth for multivariate functional data. *Advances in Data Analysis and Classification*, 8(3), 321–338.
- Martínez-Hernández, I., Genton, M. G., & González-Farías, G. (2019). Robust depth-based estimation of the functional autoregressive model. *Computational Statistics & Data Analysis*, 131, 66–79.
- Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression: A second course in statistics* (Addison-Wesley Series in behavioral science: Quantitative methods). Reading, MA: Addison-Wesley.
- Ramsay, J., & Silverman, B. (2005). Tools for exploring functional data. *Functional data analysis* (2nd ed.). 22–26. New York, NY: Springer.
- Rousseeuw, P. J., Raymaekers, J., & Hubert, M. (2018). A measure of directional outlyingness with applications to image data and video. *Journal of Computational and Graphical Statistics*, 27(2), 345–359.
- Siegel, A. F. (1983). Low median and least absolute residual analysis of two-way tables. *Journal of the American Statistical Association*, 78(382), 371–374.
- Srivastava, A., Wu, W., Kurtek, S., Klassen, E., & Marron, J. S. (2011). Registration of functional data using Fisher-Rao metric. arXiv preprint arXiv:1103.3817.
- Sun, Y., & Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2), 316–334.
- Sun, Y., & Genton, M. G. (2012a). Adjusted functional boxplots for spatio-temporal data visualization and outlier detection. *Environmetrics*, 23(1), 54–64.
- Sun, Y., & Genton, M. G. (2012b). Functional median polish. *Journal of Agricultural, Biological, and Environmental Statistics*, 17(3), 354–376.
- Sun, Y., Genton, M. G., & Nychka, D. W. (2012). Exact fast computation of band depth for large functional datasets: How quickly can one million curves be ranked? *Stat*, 1(1), 68–74.
- Tucker, J. D., Wu, W., & Srivastava, A. (2013). Generative models for functional data using phase and amplitude separation. *Computational Statistics & Data Analysis*, 61, 50–66.
- Tukey, J. W. (1970). *Exploratory data analysis* (Vol. 1, Limited Preliminary ed.). Reading, MA: Addison-Wesley.
- Tukey, J. W. (1975). *Mathematics and the picturing of data*. Proceedings of the International Congress of Mathematicians, Vancouver, 1975 (vol. 2, pp. 523–531).
- Zuo, Y. (2003). Projection-based depth functions and associated medians. *The Annals of Statistics*, 31(5), 1460–1490.

SUPPORTING INFORMATION

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

How to cite this article: Qu Z, Dai W, Genton MG. Robust functional multivariate analysis of variance with environmental applications. *Environmetrics*. 2021;32:e2641. <https://doi.org/10.1002/env.2641>