Sains Malaysiana 53(6)(2024): 1463-1476
*http://doi.org/10.17576/jsm-2024-5306-18*

# Spatial Functional Outlier Detection in Multivariate Spatial Functional Data

(Pengesanan Outlier Fungsian Reruang dalam Data Fungsian Reruang Multivariat)

NUR FATIHAH MOHD ALI[1], ROSSITA MOHAMAD YUNUS[1,*], IBRAHIM MOHAMED[1] & FARIDAH OTHMAN[2]

[1]*Institute of Mathematical Sciences, Faculty of Science, Universiti Malaya, 50603 Kuala Lumpur, Malaysia*
[2]*Department of Civil Engineering, Faculty of Engineering, Universiti Malaya, 50603 Kuala Lumpur, Malaysia*

*Received: 11 January 2024/Accepted: 23 May 2024*

ABSTRACT

Multivariate spatial functional data consists of multiple functions of time-dependent attributes observed at each spatial point. This study focuses on detecting spatial outliers in spatial functional data. Firstly, we develop a new method called Mahalanobis Distance Spatial Outlier (MDSO) to detect functional outliers in the data. The method introduces the multivariate functional Mahalanobis semi-distance and multivariate pairwise functional Mahalanobis semi-distance metrics based on the multivariate functional principal components analysis to calculate the dissimilarity between functions at each spatial point. Via simulation, we show that MDSO performs better than the other competing methods. Secondly, MDSO has been extended to detect spatial functional outliers as well. The functional outliers can now be categorized as global or/and local functional outliers. The appropriate number of neighbors and the cut-off point for the degree of isolation are determined via simulation. Finally, we demonstrate the application of the MDSO on a water quality data set obtained from Sungai Klang basin in Malaysia. The results can be used to support the authority in making better decisions on the management of the river basin or other spatial data with time-independent attributes.

Keywords: Functional Mahalanobis distance; multivariate functional data; spatial outlier; water quality

ABSTRAK

Data reruang multivariat berfungsi adalah terdiri daripada pelbagai atribut berfungsi mengikut masa yang dicerap bagi setiap titik reruang. Kajian ini mengutamakan pengesanan reruang terpencil dalam data reruang berfungsi. Pertama, kajian ini membangunkan kaedah baharu yang dikenali sebagai Jarak Mahalanobis Reruang Terpencil (JMRT) untuk mengesan fungsi terpencil dalam data. Kaedah ini memperkenalkan penganggar separa multivariat Mahalanobis berfungsi dan penganggar separa multivariat Mahalanobis berfungsi berpasangan berdasarkan analisis komponen utama multivariat berfungsi bagi menghitung perbezaan antara fungsi pada setiap titik reruang. Melalui simulasi, kajian menunjukkan bahawa prestasi JMRT lebih baik berbanding daripada kaedah lain. Kedua, kaedah JMRT dilanjutkan untuk mengesan reruang terpencil berfungsi. Fungsi terpencil yang sedia ada boleh dikategorikan kepada pencilan global dan/atau lokal berfungsi. Bilangan jiran dan titik potong bagi darjah keberasingan yang sesuai ditentukan melalui simulasi. Akhirnya, kami mengadaptasi aplikasi kaedah JMRT terhadap data kualiti air yang diambil dari lembangan Sungai Klang di Malaysia. Hasil keputusan dapat membantu pihak berwajib dalam membuat keputusan yang lebih baik untuk menguruskan lembangan sungai dan menguruskan data reruang yang bergantung terhadap masa.

Kata kunci: Data multivariat berfungsi; kualiti air; penganggar Mahalanobis berfungsi; ruang terpencil

## INTRODUCTION

The emergence of modern technology leads to a collection of large-scale data with multiple covariates stored in both space and time. The complex data set can also be transformed into a functional form and analyzed using spatial functional data analysis (SFDA). In SFDA, discrete-time observations taken at a spatial point can be quantified in the form of a function. The function is obtained through a smoothing method that fits the discrete point observations and is treated as a unique identity at a spatial point (Delicado et al. 2010). The interest in evaluating and modelling the correlated functional data has developed

naturally in many applied science disciplines when the functions were obtained for different sites (Aristizabal, Giraldo & Mateu 2019).

Multivariate functional data is an extension of univariate functional data. As with univariate functional data, multivariate functional data may also be contaminated by abnormal functions. These functions are referred to as functional outliers. This type of outlier is defined by Febrero, Galeano and González-Manteiga (2008) as curves that do not follow the same pattern as the other curves, which may represent the true underlying distribution of the data. Several methods have been developed for the detection of outliers in multivariate functional data. These include the depth measurement-based methods (Claeskens et al. 2014; Ieva & Paganoni 2013; López-Pintado et al. 2014), the visualized-based methods using graphical tools, such as bagplot (Hubert, Rousseeuw & Segaert 2015), functional outlier map (FOM) (Rousseeuw, Raymaekers & Hubert 2018) and the magnitude-shape (MS) plot (Dai & Genton 2018), and the functional Mahalanobis distance-based method (Berrendero, Bueno-Larraz & Cuevas 2020) using some generalisation of the Mahalanobis distance in Hilbert space (Galeano, Joseph & Lillo 2015). However, these methods do not consider the spatial components of the data in the development of the methods.

Spatial outliers can be distinguished into three types: global and/or local outliers. A global outlier can be defined as a spatial point having significantly different non-spatial attributes compared to most of the other points in the data set. Note that more than one global outlier may exist in any given data set. On the other hand, a local outlier is a spatial point having significantly different non-spatial attributes with respect to its closest neighbors (Haslett 1992). A global outlier might also be a local outlier when it has significantly different values of non-spatial attributes compared to its neighbors as well, referred to herein as a global and local outlier. The detection of global outliers for multivariate spatial data can be performed based on robust Mahalanobis distances, while the detection of local outliers for multivariate spatial data can be carried out based on the degree of isolation of the observation from its neighbors (Filzmoser, Ruiz-Gazen & Thomas-Agnan 2014). The degree of isolation is calculated based on pairwise Mahalanobis distances with its neighbors.

With the same motivation, we aim to propose a new method for detecting spatial functional outliers in a multivariate spatial functional dataset. The structure of multivariate spatial functional data follows closely the works of Delicado et al. (2010) and Mateu and Giraldo (2021). In the next section, we introduce the multivariate functional Mahalanobis semi-distance and the multivariate pairwise functional Mahalanobis semi-distance based on the multivariate functional principal component analysis. Hence, we may then use the distances in a new method called Mahalanobis Distance Spatial Outlier (MDSO) to detect functional outliers in the data. The method can be extended to detect spatial functional outliers, which will be further labelled as global or/and local types of outliers. The full steps of the MDSO method are given. Next, we present the simulation results to identify the $k$ nearest neighbor and the cut-off point for the degree of isolation. We compare the performance of the MDSO method to the existing functional outlier detection methods via a simulation study. Then, we apply the method to a real Malaysian water quality dataset, followed by discussion and conclusion.

## Materials and Methods

### Multivariate functional data

The notations used in Happ and Greven (2018) are closely followed in this paper. Let the multivariate functional data be a vector-valued stochastic process $X = (X_1, \ldots, X_P)'$ with $P \geq 1$ an integer. For $1 \leq p \leq P$, let $I_p$ be a compact set in R, with finite Lebesgue measure such that $X_p : I_p \longrightarrow R$ is assumed to belong $L^2(I_p)$. We denote $I := I_1 \times \ldots \times I_P$, $P$-fold Cartesian product of $I_p$. Let X be a stochastic process indexed by $t = (t_1, \ldots, t_P) \in I$ and taking values in the P-fold Cartesian product space $H := L^2(I_1) \times \ldots \times L^2(I_P)$. Let the inner product $\langle\langle \cdot, \cdot \rangle\rangle : H \times H \longrightarrow R$,

$$\langle\langle f, g \rangle\rangle := \sum_{p=1}^{P} \int_{I_p} f_p(t_p) g_p(t_p) dt_p, \quad f, g \in H.$$

Then, $H$ is a Hilbert space with respect to the scalar product $\langle\langle \cdot, \cdot \rangle\rangle$ (Happ & Greven 2018). Let $\|\cdot\|$, the norm induced by $\langle\langle \cdot, \cdot \rangle\rangle$.

We assume that $E[X(t)] := (E[X_1(t_1)], \ldots, [X_P(t_P)]) = 0, \forall t \in I$. Let $C$ denote the $P \times P$ matrix-valued covariance function which, for $s, t \in I$, is defined as $C(s,t) := E[X(s)\, X(t)']$, where the $(p,q)$th element of the matrix $C(s,t)$, for $1 \leq p, q \leq P$, is the covariance function between the $p$th and $q$th components $X : C_{p,q}(s_p, t_q) := E[X_p(s_p), X_q(t_q)] = Cov(X_p(s_p), X_q(t_q))$, $s_p \in I_p$, $t_q \in I_q$. In particular, $C_{p,q}(\cdot,\cdot)$ belongs to $L^2(I_p \times I_q)$. Let $\Gamma : H \longrightarrow H$ denotes the covariance operator of $X$ on the Hilbert space $H$, where for $f \in H$ and $t \in I$, the $q$th component of

$\Gamma f(t)$ is given by $(\Gamma f)^{(q)}(t_q) := \langle\langle C_{\cdot,q}(\cdot,t_q), f(\cdot)\rangle\rangle = \sum_{p=1}^{P} \int_{I_p} C_{p,q}(s_p, t_q) f_p(s_p) ds_p$, $t_q \in I_q$, $f \in H$.

## MULTIVARIATE KARHUNEN-LOÉVE REPRESENTATION

The representation of Karhunen-Loéve for multivariate functional data exists as $\Gamma$ has the same properties as the covariance operator in the univariate case, that is, a linear, self-adjoint and positive operator (Happ & Greven 2018). According to the theory of Hilbert-Schmidt operators, there exists a complete orthonormal basis of eigenfunctions $\{\phi_j, j = 1, 2, ...\} \subset H$ and a sequence of real numbers $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ such that $\Gamma\phi_j = \lambda_j \phi_j$ and $\phi_j \longrightarrow 0$ as $j \longrightarrow \infty$. The $\lambda_j$'s are the eigenvalues of the covariance operator $\Gamma$ and the $\phi_j$'s are the associated eigenfunctions. The multivariate version of the Karhunen-Loéve representation is $X(t) = \sum_{j=1}^{\infty} \xi_j \phi_j(t)$, $t \in I$, with zero mean random variables $\xi_j = \langle\langle X, \phi_j\rangle\rangle$ and $Cov(\xi_j, \xi_l) = \lambda_j 1_{\{j=l\}}$ (Golovkine, Klutchnikoff & Patilea 2021).

Let $J \geq 1$ and assume that the first J eigenvalues are nonzero, i.e., $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_J \geq \lambda_{J+1}$. Up to a sign, the elements of the multivariate functional principal component analysis basis are characterized by:

$\phi_1 = arg\ max_{\phi} \langle\langle \Gamma\phi, \phi\rangle\rangle$ such that $\phi = 1$,

$\phi_2 = arg\ max_{\phi} \langle\langle \Gamma\phi, \phi\rangle\rangle$ such that $\phi = 1$, and $\langle\langle \phi, \phi_1\rangle\rangle = 0$,

$\vdots$

$\phi_{J+1} = arg\ max_{\phi} \langle\langle \Gamma\phi, \phi\rangle\rangle$ such that $\phi = 1$, and $\langle\langle \phi, \phi_l\rangle\rangle = 0$, $\forall l \leq J$.

Details will be given in the next section. Then, the truncated Karhunen-Loéve expansion of the process $X$ is

$$X_{[J]}(t) = \sum_{j=1}^{J} \xi_j \phi_j(t), t \in I,\ J \geq 1; \quad (1)$$

and the truncated Karhunen-Loéve expansion of the components of $X$ is

$$X_{p_{[J_p]}}(t_p) = \sum_{j=1}^{J_p} \xi_{p,j} \phi_{p,j}(t_p), \quad t_p \in I_p,\ J_p \geq 1 \quad 1 \leq p \leq P \quad (2)$$

where $\{\phi_{p,j}, j = 1, 2, \ldots\}$ is the univariate FPCA basis associated to the covariance operator $\Gamma_p$ of $X_p$ and the scores are $\xi_{p,j} = \langle X_p, \phi_{p,j}\rangle$. Happ and Greven (2018) derived a direct relationship between the truncated representation given by Equation (2) of the single elements $X_p$ and the truncated representation given by Equation (1) of the multivariate functional data $X$.

## MULTIVARIATE FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS (MFPCA)

We now look at the theory of the multivariate functional principal component analysis using the Karhunen Loéve representation (Happ & Greven 2018). The principal component elements are, in general, not known and have to be estimated from a sample that is possibly observed on different sparse grid points. These elements are the eigenvalues $\{\lambda_j\}_{j \geq 1}$, the eigenfunctions $\{\phi_j\}_{j \geq 1}$ and the scores $\{\xi_j\}_{j \geq 1}$. Given a sample of $n$ spatial observations $X^{(1)}, \ldots, X^{(n)}$ observed at stations $s_1, \ldots, s_n$, the estimation procedure for MFPCA is shown in Figure 1.

The estimated eigenvalues and eigenfunctions are derived under the assumption of a finite sample size $n$ and a finite Karhunen Loéve representation for each univariate function $X_p$. They are relevant in practice with an appropriate choice of the truncation orders.

## THE PROPOSED MULTIVARIATE FUNCTIONAL MAHALANOBIS SEMI-DISTANCE

The functional Mahalanobis distance between a functional random variable and its corresponding mean is studied by Galeano, Joseph and Lillo (2015). Here, we introduce a new definition of multivariate functional Mahalanobis semi-distance. The multivariate functional Mahalanobis semi-distance between $X$ and mean function $\mu_X = E[X(t)]$ is denoted by

$$d_M^K(X, \mu_X) = \sqrt{\langle\langle \Gamma_K^{-\frac{1}{2}}(X - \mu_X), \Gamma_K^{-\frac{1}{2}}(X - \mu_X)\rangle\rangle}, \quad (3)$$

where $\Gamma_K^{-\frac{1}{2}}(X) = \sum_{k=1}^{K} \lambda_k^{-\frac{1}{2}} (\phi_k \otimes \phi_k(X)) = \sum_{k=1}^{K} \lambda_k^{-\frac{1}{2}} \langle\langle \phi_k, X\rangle\rangle (\phi_k)$ is a regularized square root inverse operator for a given threshold $K$ and $X$ is a function in the range of $\Gamma_K$. We then have the multivariate functional Mahalanobis semi-distance in terms of the multivariate functional principal component scores of $X$ as given by

$$d_M^K(X, \mu_X) = \sqrt{\sum_{k=1}^{K} \lambda_k^{-1} \langle\langle X - \mu_X, \phi_k\rangle\rangle^2}. \quad (4)$$

Given the regularized square root inverse operator, $\Gamma_K^{-\frac{1}{2}}(X) = \sum_{k=1}^{K} \lambda_k^{-\frac{1}{2}} (\phi_k \otimes \phi_k(X))$ and the multivariate

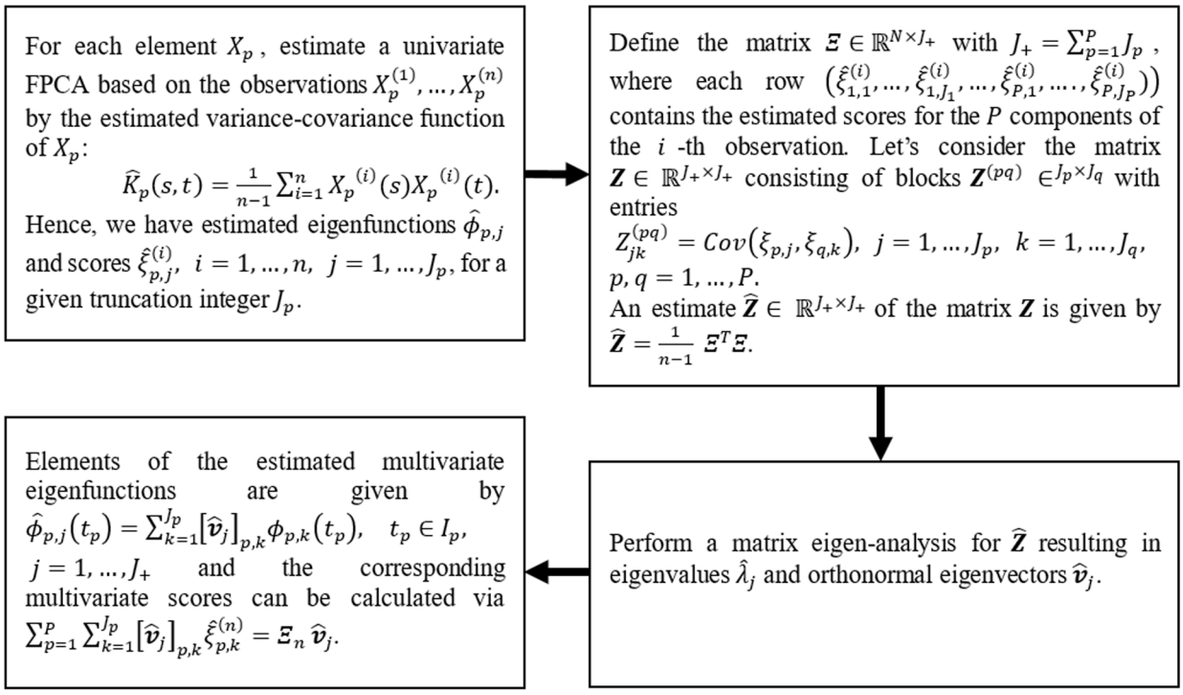


FIGURE 1. Flowchart of MFPCA

Karhunen Loéve representation $X(t) = \mu_X + \sum_{j=1}^{\infty}\xi_j\phi_j(t)$, $t \in I$, with random variables $\xi_j = \langle\langle X - \mu_X, \phi_j\rangle\rangle$ and $Cov(\xi_j, \xi_l) = \lambda_l 1_{\{j=l\}}$, then we can rewrite Equation (4) as

$$d_M^K(X, \mu_X) = \sqrt{\langle\langle \Gamma_K^{-\frac{1}{2}}(X - \mu_X), \Gamma_K^{-\frac{1}{2}}(X - \mu_X)\rangle\rangle}$$

$$= \sqrt{\langle\langle \sum_{k=1}^{K} \lambda_k^{-\frac{1}{2}}(\phi_k \otimes \phi_k(X - \mu_X)), \sum_{k=1}^{K} \lambda_k^{-\frac{1}{2}}(\phi_k \otimes \phi_k(X - \mu_X))\rangle\rangle}$$

With the property of cross product $a \otimes a\,(c) = \langle a, c\rangle(a)$ and $X - \mu_X = \sum_{j=1}^{\infty}\xi_j\phi_j$, we have

$$d_M^K(X, \mu_X) =$$

$$\sqrt{\langle\langle \sum_{k=1}^{K} \lambda_k^{-\frac{1}{2}}\langle\langle \phi_k, \sum_{j=1}^{\infty} \xi_j\phi_j\rangle\rangle(\phi_k), \sum_{k=1}^{K} \lambda_k^{-\frac{1}{2}}\langle\langle \phi_k, \sum_{j=1}^{\infty} \xi_j\phi_j\rangle\rangle(\phi_k)\rangle\rangle}$$

$$=$$

$$\sqrt{\langle\langle \sum_{k=1}^{K} \lambda_k^{-\frac{1}{2}}\sum_{j=1}^{\infty} \xi_j\langle\langle \phi_k, \phi_j\rangle\rangle(\phi_k), \sum_{k=1}^{K} \lambda_k^{-\frac{1}{2}}\sum_{j=1}^{\infty} \xi_j\langle\langle \phi_k, \phi_j\rangle\rangle(\phi_k)\rangle\rangle}$$

$$= \sqrt{\sum_{k=1}^{K} \lambda_k^{-1}\langle\langle \sum_{j=1}^{\infty} \xi_j\langle\langle \phi_k, \phi_j\rangle\rangle(\phi_k), \sum_{j=1}^{\infty} \xi_j\langle\langle \phi_k, \phi_j\rangle\rangle(\phi_k)\rangle\rangle}.$$

Since $\phi_k$ being orthonormal eigenfunctions, then, we can further express the formula as

$$d_M^K(X, \mu_X) = \sqrt{\sum_{k=1}^{K} \lambda_k^{-1}\sum_{j=1}^{\infty} \xi_j^2\langle\langle \phi_k, \phi_j\rangle\rangle^2}$$

$$= \sqrt{\sum_{k=1}^{K} \lambda_k^{-1}\xi_k{}^2}$$

$$= \sqrt{\sum_{k=1}^{K} \frac{1}{\lambda_k}\langle\langle X - \mu_X, \phi_k\rangle\rangle^2}.$$

Next, we extend the definition of multivariate functional Mahalanobis semi-distance to the pairwise multivariate functional Mahalanobis semi-distance. The pairwise multivariate functional Mahalanobis semi-distance measures the distance between two identically distributed functional random variables, $X^{(g)}$ and $X^{(h)}$ observed at $s_g$ and $s_h$ stations, respectively where $g \neq h$ is given by

$$d_M^K\left(X^{(g)}, X^{(h)}\right) = \sqrt{\langle\langle \Gamma_K^{-\frac{1}{2}}(X^{(g)} - X^{(h)}), \Gamma_K^{-\frac{1}{2}}(X^{(g)} - X^{(h)})\rangle\rangle}$$

$$= \sqrt{\langle\langle \sum_{k=1}^{K} \lambda_k^{-\frac{1}{2}}(\phi_k \otimes \phi_k(X^{(g)} - X^{(h)})), \sum_{k=1}^{K} \lambda_k^{-\frac{1}{2}}(\phi_k \otimes \phi_k(X^{(g)} - X^{(h)}))\rangle\rangle}.$$

With the property of cross product $a \otimes a\,(c) = \langle a,c\rangle(a)$ and $X^{(g)} - X^{(h)}) = \sum_{j=1}^{\infty}(\xi_{gk} - \xi_{hk})\,\phi_j$, we have

$$d_M^K\left(X^{(g)}, X^{(h)}\right) =$$

$$\sqrt{\langle\langle \sum_{k=1}^{K} \lambda_k^{-\frac{1}{2}}\langle\langle \phi_k, X^{(g)} - X^{(h)}\rangle\rangle(\phi_k), \sum_{k=1}^{K} \lambda_k^{-\frac{1}{2}}\langle\langle \phi_k, X^{(g)} - X^{(h)}\rangle\rangle(\phi_k)\rangle\rangle}$$

$$\sqrt{\sum_{k=1}^{K} \lambda_k^{-1}\langle\langle \langle\langle \phi_k, \sum_{j=1}^{\infty} (\xi_{gk} - \xi_{hk})\phi_j\rangle\rangle(\phi_k), \langle\langle \phi_k, \sum_{j=1}^{\infty} (\xi_{gk} - \xi_{hk})\phi_j\rangle\rangle(\phi_k)\rangle\rangle}$$

Since $\phi_k$ being orthonormal eigenfunctions, we can show that

$$d_M^K\left(X^{(g)}, X^{(h)}\right) = \sqrt{\sum_{k=1}^{K} \lambda_k^{-1}\langle\langle (\xi_{gk} - \xi_{hk})\phi_k, (\xi_{gk} - \xi_{hk})\phi_k\rangle\rangle}$$
$$= \sqrt{\sum_{k=1}^{K} \frac{1}{\lambda_k}(\xi_{gk} - \xi_{hk})^2}\,.$$

Thus, the pairwise multivariate functional Mahalanobis semi-distance can be written in the form of standardized multivariate functional principal component scores of $X^{(g)}$ and $X^{(h)}$ as

$$d_M^K\left(X^{(g)}, X^{(h)}\right) = \sqrt{\sum_{k=1}^{K} \lambda_k^{-1}(\xi_{g,k} - \xi_{h,k})^2}\,, \quad (5)$$

where $\xi_{i,k} = \langle\langle X^{(i)} - \mu_{X}, \phi_k\rangle\rangle$, for $i = 1, \ldots, n$.

$$= \sum_{p=1}^{P} \int (X_{i,p}(t_p) - \mu_{X_{i,p}}), \phi_{k,p}(t_p)\, dt_p$$

$$= \sum_{p=1}^{P} \int_{I_p} (X_{i,p}(t_p) - \mu_{X_{i,p}})\, \phi_{k,p}(t_p)\, dt_p$$

Following Berrendero, Bueno-Larraz and Cuevas (2020) and Galeano, Joseph and Lillo (2015), the distribution of the squared functional Mahalanobis distance, $d_M^{K\,2}(X, \mu_X)$ for a Gaussian process $X$ follows chi-square with $K$ degrees of freedom, $\chi_K^2$.

*Proposition 1 Let $X^{(1)}, \ldots, X^{(n)}$ be multivariate functional centred random variables on a set of locations $s_1, \ldots, s_n$ in a spatial domain $D \subset R^2$, and the scores $\hat{\xi}_i = \left(\hat{\xi}_{i,1}, \ldots, \hat{\xi}_{i,K}\right)'$ be independent and identically variables in K dimension following Gaussian vector. Then, the conditional distribution of the pairwise squared multivariate functional Mahalanobis semi-distance $d_M^{K\,2}(X^{(g)}, X^{(h)})$, $g, h \in \{1, 2, \ldots, n\}$ given $X^{(g)}$ given follows a non-central chi-square distribution with K degrees of freedom and the non-centrality parameter $d_M^{K\,2}(X^{(g)}, \mu_X)$, the squared functional Mahalanobis semi-distance.*

Proof

Consider the pairwise squared multivariate functional Mahalanobis semi-distance $d_M^{K\,2}(X^{(g)}, X^{(h)})$ when $X^{(g)} = x$ with $x = (x_1, \ldots, x_p)' \in R^p$ in the infinite dimensional space $L^2(I_p)$ and let $y = \sum_{k=1}^{K} \lambda_k^{-1}\langle\langle x - \mu_x, \phi_k\rangle\rangle$ truncated on $K \in N$, $\lambda_k$ is the eigenvalue and $\phi_k$ is the eigenfunction of covariance operator $\Gamma_K$. Then the pairwise squared functional Mahalanobis distance,

$$d_{FM}^{K\,2}\left(X^{(g)}, x\right) = \sum_{k=1}^{K} \frac{1}{\lambda_k}\langle\langle X^{(g)} - x, \phi_k\rangle\rangle^2$$

$$= \sum_{k=1}^{K} \frac{1}{\lambda_k}\left(\int_{I_p} (X_g(t) - x(t))\phi_k(t)\, dt\right)^2$$

$$= \sum_{k=1}^{K} \frac{1}{\lambda_k}\left(\sum_{j=1}^{\infty} (\xi_{j,g} - \xi_{j,x})\langle\langle \phi_j, \phi_k\rangle\rangle\right)^2$$

$$= \sum_{k=1}^{K} \lambda_k^{-1}(\xi_{k,g} - \xi_{k,x})^2,$$

where $\xi_{k,x}$ is the scores of $x$ and $\lambda_k^{-\frac{1}{2}}(\xi_{k,g} - \xi_{k,x})$ are independent random variables with mean 0 and variance 1. Thus, $d_M^{K\,2}(X^{(g)}, x)$follow a non-central chi-square distribution with $K$ degrees of freedom and non-centrality parameter $d_M^{K\,2}(X^{(g)}, \mu_X)$.

We note that the formulation of Equations (4) and (5) reduces to the Mahalanobis distance formulation on the selected MFPCA scores.

## THE PROPOSED MDSO METHOD

This subsection proposes a spatial functional outlier detection method using the distance defined in the previous section. Consider a multivariate spatial functional data consists of p number of non-spatial attributes that are observed on n spatial observations such that $X^{(1)}, \ldots, X^{(n)}$ at stations $s_1, \ldots, s_n$, respectively. We first applying an MFPCA to the multivariate functional data. We then use the functional principal scores to detect global and/or local functional outliers. The method for detecting global functional outliers is based on the multivariate functional Mahalanobis distance given by Equation (4), where $\lambda_K$ are the eigenvalues or the measure of the variation of $K$ principal components and the scores $\hat{\xi}_i = (\hat{\xi}_{i,1}, \ldots, \hat{\xi}_{i,K})'$ be independent variables. The global functional outliers are determined when the values of the functional Mahalanobis distance given by Equation (4) are larger than the cut-off value. Here the cut-off value is the square root of 97.5% quantile of the chi-square distribution with $K$ degrees of freedom $\sqrt{\chi^2_{K;0.975}}$.

As for the detection of local functional outliers,

firstly, the Euclidean distances between station coordinates are calculated to determine the $k$ nearest neighbors. Then, the pairwise multivariate functional Mahalanobis semi-distance between a station and each of the $k$ nearest neighbors are calculated using Equation 5. The distances are then sorted in ascending order. According to Filzmoser, Ruiz-Gazen and Thomas-Agnan (2014), the degree of isolation of a station implies that the attributes at the station are very different from most of its neighbors. If only the next nearest neighbor is considered, then it would be bias because just by chance the next nearest neighbor's attributes could be close but a third neighbor might be far away. Thus, $\beta$ is denoted as a fraction and $\lceil k \cdot \beta \rceil$ is the number of neighbors of a station that can be similar to the station. Therefore, $X^{(\lceil k \cdot \beta \rceil)}$ can be understood as the functional observations of the next nearest neighbor with index$\lceil k \cdot \beta \rceil$. Then, the degree of isolation of a station, $\alpha(i)$-quantile is computed by

$$\chi^2_{K;\alpha(i)}\left(d_M^{K\,2}\left(X^{(i)},\mu_X\right)\right) = d_M^{K\,2}\left(X^{(i)},X^{(\lceil k\cdot\beta\rceil)}\right)^2 \text{ for } i = 1,\ldots,n. \quad (6)$$

The pairwise squared multivariate functional Mahalanobis semi-distance on the right-hand side in Equation (6) is a non-central chi-square distribution with $K$ degrees of freedom. The non-centrality parameter of the squared multivariate functional Mahalanobis semi-distance is represented on the left-hand side of the equation. The cut-off point is determined by $\beta$-value. If $\alpha(i)$ is significantly larger than $\beta$, then station i considered as a local functional outlier.

The proposed algorithm of the detection of spatial functional outliers is presented as follows. Given a multivariate functional data with known spatial location,

**Step 1.** Set a number of $k$ nearest neighbors and a fraction of neighbors, $\beta$.

**Step 2.** Perform MFPCA on data to obtain the eigenvalues and functional principal scores for each station.

**Step 3.** Compute $d_M^K(X,\mu_X) = \sqrt{\sum_{k=1}^{K} \lambda_k^{-1}\xi_k^{\,2}}$. If $d_M^K \geq \sqrt{\chi^2_{K;0.975}}$, station i is a global functional outlier.

**Step 4.** Calculate the Euclidean distance between station coordinates and considers only $k$ nearest neighbors.

**Step 5.** Compute the pairwise multivariate functional Mahalanobis semi-distance between the station and each of the neighbors. Sort the distances in ascending order.

**Step 6.** Calculate the next nearest neighbor index $\lceil k \cdot \beta \rceil$ and determine the pairwise multivariate functional Mahalanobis semi-distance.

**Step 7.** Then, compute the degree of isolation for each spatial point by using $\chi^2_{K;\alpha(i)}\left(d_M^{K\,2}\left(X^{(i)},\mu_X\right)\right) = d_M^{K\,2}\left(X^{(i)},X^{(\lceil k\cdot\beta\rceil)}\right)^2$ for $i = 1, \ldots, n$.

**Step 8.** Sort the value of the degree of isolation. Then,

(i) if a global outlier identified in step 3 has the degree of isolation significantly larger than $\beta$, then we classify the observation as a local and global functional outlier.
(ii) if other observation has the degree of isolation significantly larger than $\beta$, then we classify the observation as local functional outlier.
(iii) otherwise, the observation is regular observation.

## Results and Discussion

The performance of the proposed method was studied via simulation. For this purpose, we generate the multivariate spatial functional data based on a truncated Karhunen-Loéve representation of functional data (Happ & Greven 2018) with the spatial covariance that corresponds to the given distance. The result from the proposed method is then compared to that of the other functional outlier detection methods in terms of their capability to detect functional outliers in the data.

### DATA SIMULATION

The simulation data set consists of 30 spatial points or stations with the $XY$ coordinates. The coordinates are generated from a random uniform distribution. The Euclidean distance between stations is calculated to generate the spatial covariance matrix for the data. Thus, the spatial covariance matrix is defined according to the exponential model $C(h) = 2\ exp\ (-\frac{h}{100})$ with $h = \| s_i - s_j \|$, $i, j = 1,\ldots, 30$.

Next, the bivariate functional data samples with 50 time points at each station are generated. The data is modelled by the Karhunen-Loéve representation of a function $X^{(i)} = (X_1^{(i)}, X_2^{(i)})$ truncated at $J$ given by

$$X^{(i)}(t) = \mu(t) + \sum_{j=1}^{J} \xi_{i,j}\,\phi_j(t), \quad i = 1,\ldots,N, t = (t_1,t_2) \in I_1 \times I_2 \quad (7)$$

with a zero multivariate mean function $\mu(t)$ and multivariate eigenfunctions $\phi_j$, $j = 1, \ldots, J$. The

individual scores $\xi_{i,j} = \langle\langle X^{(i)}, \phi_j \rangle\rangle$ are realizations of random variables $\xi_j$ with expected value is zero and the variance is $\lambda_j$ with eigenvalues $\lambda_j > 0$. For the eigenfunctions, the scores are generated through Fourier basis function through $\xi_{i,j} \sim N(0,\lambda_j)$. For the eigenvalues $\lambda_j$, we choose a linear $\lambda_j = (J\text{-}j+1)/J$. Then, for bivariate data, the eigenfunctions are calculated based on tensor products of univariate functional data. Thus, for 30 functions on $I = [1,50]\times[1,50]$, the eigenfunctions are calculated as tensor products of $J_1$=10 eigenfunctions of the Fourier basis function on [1,50] and $J_2$ = 10 eigenfunctions of the Fourier basis function on [1,50]. The generated samples are obtained using MFPCA in *R* package.

For the multivariate spatial functional data, the generated multivariate functional data are obtained by multiply the bivariate functional data with the spatial covariance coefficent. Thus, the new data are spatially correlated which depends on the distance between the spatial points. Next, a random uniform value is added to the original function of a randomly chosen spatial point to introduce a global functional outlier in the data. As a result, the value of the functional Mahalanobis distance of this spatial point is greater than the critical chi-square value for the degree of freedom $K = 2$ with a critical alpha value of 0.025. Figure 2 shows that spatial point 3 is outside the ellipse since it is a global functional outlier but its corresponding neighbors are far inside the ellipse. Thus, spatial point 3 may also be a global and/or local functional outlier.

Meanwhile, a local functional outlier was set-up by randomly selecting a spatial point as the targeted local functional outlier and determine the nearest neighbors of the point on the XY-coordinates. Then, the function is added by the average value of the neighboring functions. So that, the function is different from its neighboring spatial points and the value of the functional Mahalanobis distance of the spatial point must be within the critical chi-square value for degree of freedom, $K = 2$ at a critical alpha value of 0.025. Another consideration should be noted is that the local functional outlier and the neighboring stations should be different from the global functional outlier in order to correctly measure the spatial functional outlier detection performance.

## SIMULATION RESULTS

For the simulations, the experiment was repeated 200 times for different sizes of neighbors, $k$ = 5, 10, 20, 30 and $\beta$ = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5. We introduce 10% and 20% contamination of outliers in the data.

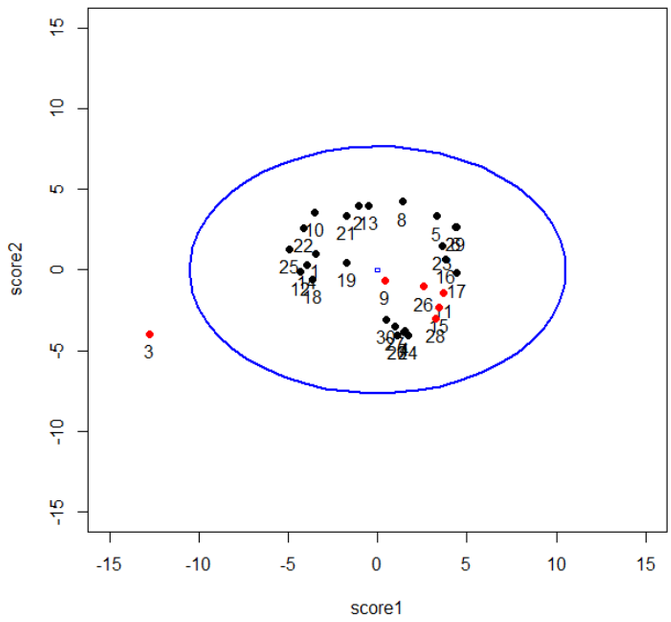


FIGURE 2. Scores plot of bivariate spatial functional data

The performance of the MDSO method is measured by calculating the misclassification error rate and area under the curve (AUC) values as adopted in previous studies (López-Pintado et al., 2014; Ojo, Fernández Anta & Lillo 2019; Sun & Genton 2011). The best results should not only detect the outliers, but also avoid misclassifying good observations as outliers. The results of the simulation are presented in Tables 1 and 2 for 10% and 20% contamination, respectively.

Overall, the method performs well when $\beta = 0.3$ where the mean AUC is consistently high for each number of neighbors $k$. The mean AUC value is high for $\beta = 0.3$ as fewer regular spatial points are misclassified as outliers. Besides, for smaller values of proportion of neighborhood $\beta = 0.05$, the mean AUC value is reduced because many regular spatial points are misclassified as outliers. Moreover, for the high proportion of neighborhood, the mean AUC values start to decrease due to many true outliers are unidentified.

The appropriate number of neighbors, $k$ depends on the sample size of the data. From both Tables 1 and 2, when the size of the data is 30 and 50, a high mean AUC value is observed for $k = 10$, while when the size is 100 and 200, a high mean AUC value is recorded for $k = 30$.

Thus, for larger sample sizes, we require larger k to obtain good outlier detection. However, the mean AUC value decreases as many neighbors are considered. This is due to the fact that, when larger k is considered, there is a higher possibility that true outliers will also be chosen among the neighbors. Hence, the corresponding value of degree of isolation will be lower and the true outliers fail to be identified. In addition, the mean AUC values are slightly higher when more percentage of outliers are added in the data and as expected, the higher the sample sizes considered, the larger the mean AUC values are.

TABLE 1. Mean AUC for different $k = 5, 10, 20, 30$, $\beta = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6$ and for 10% outliers contaminated within the data

| Sample size | $k$ | $\beta$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| 30 | 5 | 0.381 | 0.589 | 0.740 | 0.745 | 0.615 | 0.513 |
| | 10 | 0.455 | 0.554 | 0.480 | 0.750 | 0.690 | 0.618 |
| | 20 | 0.530 | 0.461 | 0.369 | 0.683 | 0.596 | 0.609 |
| | 30 | - | - | - | - | - | - |
| 50 | 5 | 0.563 | 0.695 | 0.698 | 0.802 | 0.777 | 0.523 |
| | 10 | 0.665 | 0.713 | 0.754 | 0.840 | 0.788 | 0.496 |
| | 20 | 0.610 | 0.793 | 0.801 | 0.829 | 0.795 | 0.534 |
| | 30 | 0.642 | 0.694 | 0.773 | 0.833 | 0.830 | 0.810 |
| 100 | 5 | 0.828 | 0.856 | 0.856 | 0.887 | 0.839 | 0.704 |
| | 10 | 0.815 | 0.820 | 0.886 | 0.902 | 0.840 | 0.668 |
| | 20 | 0.680 | 0.842 | 0.902 | 0.908 | 0.906 | 0.897 |
| | 30 | 0.735 | 0.856 | 0.906 | 0.909 | 0.907 | 0.903 |
| 200 | 5 | 0.906 | 0.913 | 0.913 | 0.947 | 0.909 | 0.797 |
| | 10 | 0.886 | 0.889 | 0.930 | 0.948 | 0.923 | 0.839 |
| | 20 | 0.752 | 0.885 | 0.943 | 0.950 | 0.912 | 0.782 |
| | 30 | 0.858 | 0.899 | 0.947 | 0.951 | 0.897 | 0.767 |

TABLE 2. Mean AUC for different $k$ =5, 10, 20, 30, $\beta$ = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6 and for 20% outliers contaminated within the data

| Sample size | $k$ | $\beta$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| 30 | 5 | 0.389 | 0.545 | 0.687 | 0.837 | 0.779 | 0.591 |
| | 10 | 0.598 | 0.617 | 0.680 | 0.817 | 0.784 | 0.742 |
| | 20 | 0.515 | 0.627 | 0.799 | 0.815 | 0.805 | 0.561 |
| | 30 | - | - | - | - | - | - |
| 50 | 5 | 0.672 | 0.769 | 0.793 | 0.891 | 0.877 | 0.631 |
| | 10 | 0.628 | 0.681 | 0.837 | 0.894 | 0.871 | 0.731 |
| | 20 | 0.618 | 0.711 | 0.854 | 0.889 | 0.860 | 0.708 |
| | 30 | 0.581 | 0.688 | 0.810 | 0.842 | 0.803 | 0.515 |
| 100 | 5 | 0.807 | 0.843 | 0.848 | 0.904 | 0.881 | 0.654 |
| | 10 | 0.718 | 0.750 | 0.860 | 0.910 | 0.903 | 0.663 |
| | 20 | 0.630 | 0.695 | 0.892 | 0.930 | 0.906 | 0.748 |
| | 30 | 0.653 | 0.711 | 0.899 | 0.946 | 0.902 | 0.804 |
| 200 | 5 | 0.871 | 0.888 | 0.888 | 0.940 | 0.928 | 0.814 |
| | 10 | 0.794 | 0.814 | 0.906 | 0.949 | 0.944 | 0.850 |
| | 20 | 0.613 | 0.734 | 0.913 | 0.967 | 0.948 | 0.861 |
| | 30 | 0.616 | 0.696 | 0.918 | 0.971 | 0.941 | 0.856 |

## COMPARISON OF METHOD

The performance of the MDSO method is compared with the existing methods of outlier detection for multivariate functional data. We review the following existing outlier detection methods, namely, Magnitude-Shape Plot or MS-Plot (Dai & Genton 2018), Functional Outlier Map or FOM (Rousseeuw, Raymaekers & Hubert 2018), Weighted Modified Band Depth or WMBD (Ieva & Paganoni 2013), Multivariate Outliergram or MulOut (Arribas-Gil & Romo 2014) and Modified simplicial band depth or MSBD (López-Pintado et al. 2014).

The results in Table 3 show that the performance of the proposed detection method, MDSO is generally better than the other existing methods. For data with 10% contamination, MDSO, MS-Plot and FOM detect all true outliers and only MS-Plot shows 27% false detection whereas the other two methods recorded no false detection. Thus, the accuracy for MDSO and the MS-Plot are 100% while the accuracy for MS-Plot is only 75%. The WMBD, MulOut, and MSBD performance are also good with the accuracy is more than 50% but the TPR is too low and the FPR is considerably high.

For the case of a higher rate of contamination (15% and 20% of outliers in the data), the same pattern of results is observed, but with lower performance measures for all methods. In addition, the accuracy result for MDSO is still the highest compared to the other methods, indicating the superiority of the proposed method. Some methods, such as the FOM, are comparable in performance to the MDSO in detecting functional outliers. However, unlike MDSO, they cannot distinguish the types of outliers in the simulation study.

## APPLICATION TO REAL DATA

We apply the proposed spatial functional outlier detection method to a real data set, the water quality

of Sungai Klang data set. The data contains monthly observations of seven water quality parameters for $s = 35$ stations, averaged over the years 2013 to 2016, provided by the Department of Environment Malaysia. The water quality monitoring stations are located within the states of Selangor and Kuala Lumpur in Malaysia (Figure 3). The data consists of several water quality parameters, which are dissolve oxygen (DO), bio-chemical oxygen demand (BOD), chemical oxygen demand (COD), suspended solids (SS), ammoniacal nitrogen ($NH_3NL$), temperature, and pH. These parameters are important for assessing the quality of river water. In addition, the $XY$ coordinates of these stations are also recorded in the data. Most of the monitoring stations are located in the middle stretch of Sungai Klang basin, which indicates the importance of preserving good water quality in the area.

We apply the MDSO method by selecting the number of neighbors, $k = 10$ and the proportion of neighbors, $\beta = 0.3$. The MDSO method identified Stations 2 and 6 as the global functional outliers as the stations have high multivariate functional Mahalanobis semi-distances which exceed the cut-off point, $\sqrt{\chi^2_{3;0.975}}$ as shown by Figure 4.

TABLE 3. Mean and standard deviation (in parentheses) of TPR, FPR and accuracy of outlier detections for each outlier detection methods with different contamination rates

| Method | TPR | FPR | Accuracy |
|---|---|---|---|
| 10% contamination | | | |
| MDSO | 1.00 (0.00) | 0.00(0.00) | 1.00 (0.00) |
| MS-plot | 1.00 (0.00) | 0.27(0.04) | 0.75 (0.04) |
| FOM | 1.00 (0.00) | 0.00(0.00) | 1.00 (0.00) |
| WMBD | 0.20 (0.00) | 0.00(0.00) | 0.92 (0.00) |
| MulOut | 0.00 (0.00) | 0.09 (0.02) | 0.81 (0.02) |
| MSBD | 0.39 (0.24) | 0.28 (0.03) | 0.68 (0.05) |
| 15% contamination | | | |
| MDSO | 0.979 (0.05) | 0.000 (0.00) | 0.997 (0.01) |
| MS-plot | 1.000 (0.00) | 0.247 (0.05) | 0.788 (0.04) |
| FOM | 0.993 (0.03) | 0.003 (0.02) | 0.996 (0.01) |
| WMBD | 0.143 (0.00) | 0.000 (0.00) | 0.88 (0.00) |
| MulOut | 0.000 (0.00) | 0.094 (0.02) | 0.779 (0.02) |
| MSBD | 0.236 (0.12) | 0.310 (0.02) | 0.626 (0.03) |
| 20% contamiantion | | | |
| MDSO | 0.890 (0.14) | 0.000 (0.00) | 0.978 (0.03) |
| MS-plot | 1.000 (0.00) | 0.181 (0.07) | 0.855 (0.05) |
| FOM | 0.840 (0.30) | 0.000 (0.00) | 0.968 (0.06) |
| WMBD | 0.100 (0.00) | 0.000 (0.00) | 0.820 (0.00) |
| MulOut | 0.000 (0.00) | 0.109 (0.01) | 0.713 (0.01) |
| MSBD | 0.250 (0.10) | 0.304 (0.03) | 0.607 (0.04) |

As for local functional outliers, the MDSO method gives the degree of isolation as shown in Figure 5. Only one station, which is Station 6, has a high degree of isolation that exceeds the dashed line at 0.3. Thus, Station 6 is identified as a local functional outlier. In addition, Station 6 is also identified as a global functional outlier and hence, this station is identified as a global and local functional outlier.

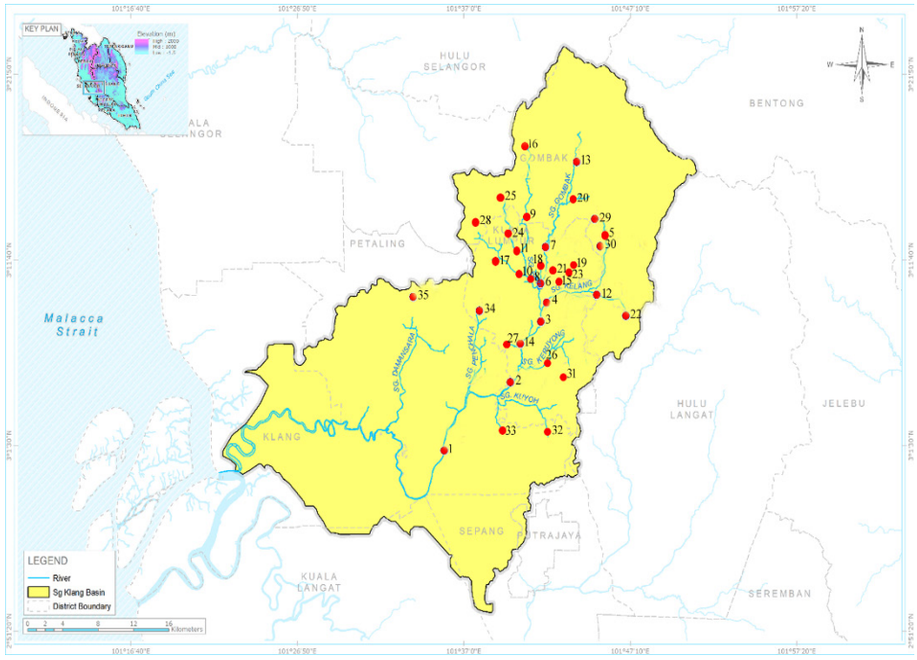


FIGURE 3. Map of 35 stations in Sungai Klang basin

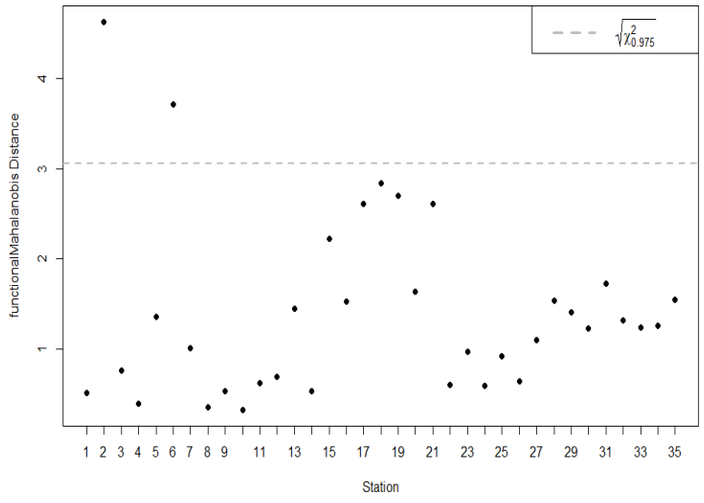


FIGURE 4. Multivariate functional Mahalanobis semi-distance

## DISCUSSION

We consider Stations 2 and 6 for a detailed description of the results. Figure 6 shows both stations that are detected as global functional outliers have remarkably high TSS values at the beginning and end of the time study.

In Figure 7, for every water quality parameter, the functional curve for Station 6 (red) is slightly different compared to its neighbors (blue), even though Station 6 (red) is geographically close to its neighbors (blue) on the map. The DO values of Station 6 are slightly higher than its neighbors in almost cases. For example, from the beginning of the time study until the middle of 2013 and from the beginning of the year 2014 until the middle of 2014. The DO values of Station 6 were the highest from the end of 2015 until the beginning of 2016.

In addition, the corresponding function of TSS for Station 6 has a significant difference in magnitude and shape compared to its neighboring stations. As shown in Figure 7, from January 2013 until April 2013, the TSS function (red) lies farther above 100 mg/L. Then, the function steadily fluctuated above its neighbors (blue) until the middle of 2014. Next, even though the function (red) is within the range of its neighbors (blue) until the end of the study, the difference in shape of the function (red) is clear, especially at the end of 2016.

Based on the comprehensive analysis of Stations 2 and 6, a valuable insights regarding the river water quality can be obtained. By identifying the functional outliers and highlighting the significant deviations of the decremental water quality parameters, this analysis enables the authorities to gain critical information for understanding the temporal and spatial variations in water quality. This is particularly important, especially in the densely populated areas or industrial zones. Moreover, these information enables proactive initiative to reduce potential environmental risks and ensure sustainable and effective management of water resources in the region. In addition, the information can be used to aid the authorities in making better decisions on the management of the river basin.

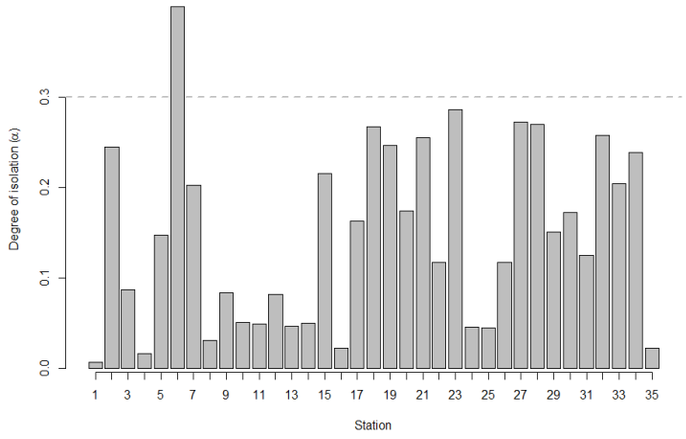


FIGURE 5. Degree of isolation for each station

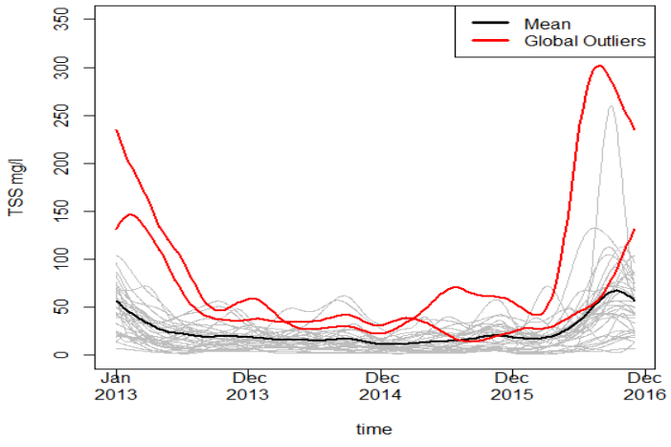


FIGURE 6. The smoothed functional data for TSS mg/L

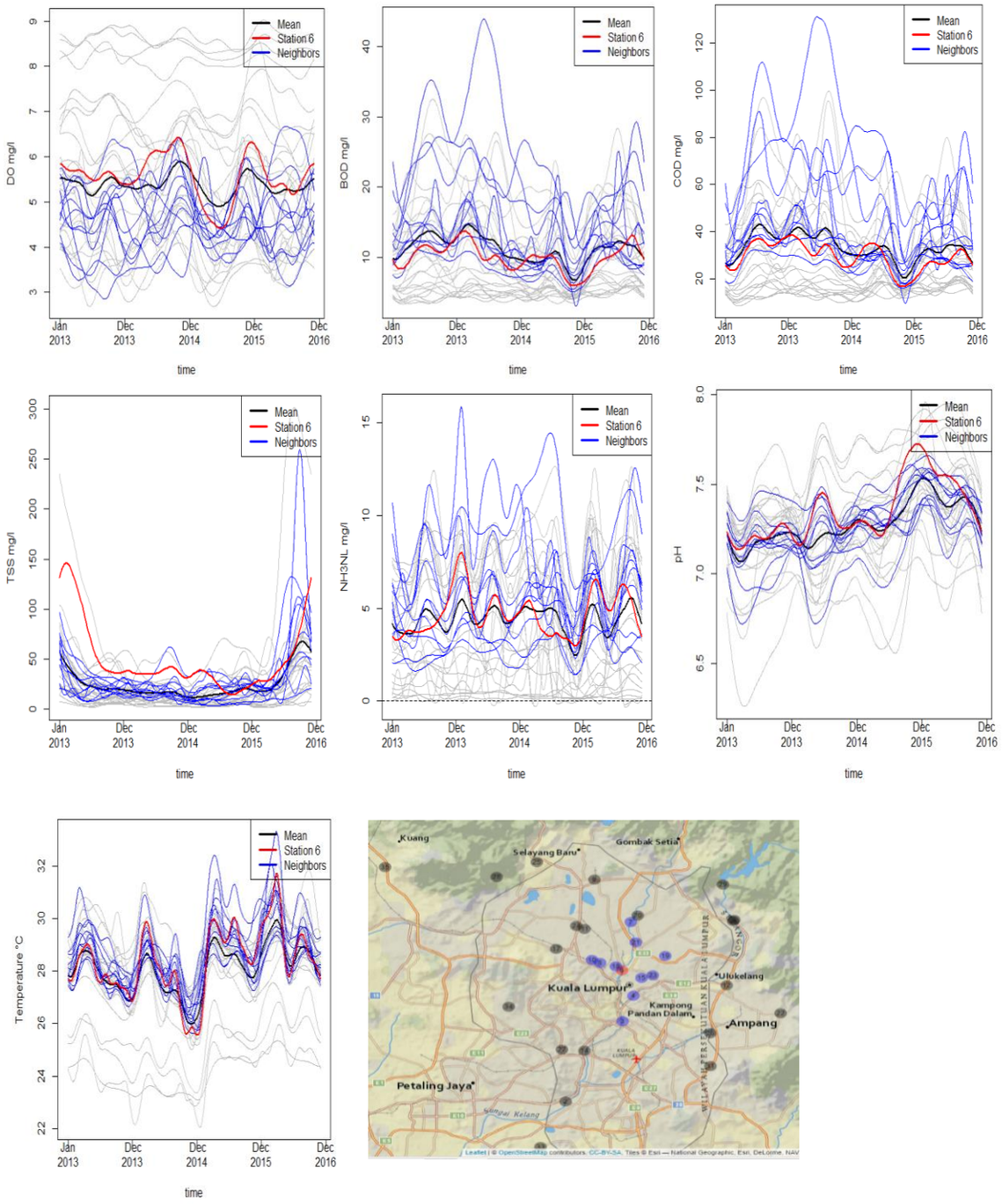


FIGURE 7. The multivariate functional data and the map of the stations for the local functional outliers and the corresponding neighbors

## CONCLUSION

In this paper, we propose a new detection method for identifying spatial outliers in multivariate functional datasets. We introduce the multivariate pairwise functional Mahalanobis semi-distance metrics based on the multivariate functional principal components analysis to calculate the dissimilarity between functions of attributes at each spatial point. Then, we develop a new method

called Mahalanobis Distance Spatial Outlier (MDSO) to detect spatial functional outliers. In the simulation study, the MDSO method outperforms existing methods by accurately detecting multivariate spatial functional outlier and global and/or local functional outliers. The study shows that the performance of the outlier detection method is good when the fraction of neighbors, $\beta$ is equal to 0.3 and the number of nearest neighbors, k is not chosen too large; in this experiment, k is equal to 10. For the real data application, we consider water quality data of Sungai Klang basin. We detect one station as the global and local functional outlier using the MDSO. The water quality parameter function at this station diverges significantly in both magnitude and shape from those of its neighbors. By identifying the outliers, we highlight the necessity for targeted intervention and management strategies to be implemented in this area. In conclusion, this study contributes not only to the development of outlier detection methods in functional data but also provides the need to understand the effect of the types of outliers in multivariate spatial functional data.

## REFERENCES


Aristizabal, J.P., Giraldo, R. & Mateu, J. 2019. Analysis of variance for spatially correlated functional data: Application to brain data. *Spatial Statistics* 32: 100381.

Arribas-Gil, A. & Romo, J. 2014. Shape outlier detection and visualization for functional data: The outliergram. *Biostatistics* 15(4): 603-619.

Berrendero, J.R., Bueno-Larraz, B. & Cuevas, A. 2020. On Mahalanobis distance in functional settings. *The Journal of Machine Learning Research* 21(1): 288-320.

Claeskens, G., Hubert, M., Slaets, L. & Vakili, K. 2014. Multivariate functional halfspace depth. *Journal of the American Statistical Association* 109(505): 411-423.

Dai, W. & Genton, M.G. 2018. Multivariate functional data visualization and outlier detection. *Journal of Computational and Graphical Statistics* 27(4): 923-934.

Delicado, P., Giraldo, R., Comas, C. & Mateu, J. 2010. Statistics for spatial functional data: Some recent contributions. *Environmetrics: The Official Journal of the International Environmetrics Society* 21(3-4): 224-239.

Febrero, M., Galeano, P. & González-Manteiga, W. 2008. Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels. *Environmetrics: The Official Journal of the International Environmetrics Society* 19(4): 331-345.

Filzmoser, P., Ruiz-Gazen, A. & Thomas-Agnan, C. 2014. Identification of local multivariate outliers. *Statistical Papers* 55: 29-47.

Galeano, P., Joseph, E. & Lillo, R.E. 2015. The Mahalanobis distance for functional data with applications to classification. *Technometrics* 57(2): 281-291.

Golovkine, S., Klutchnikoff, N. & Patilea, V. 2021. Adaptive optimal estimation of irregular mean and covariance functions. *arXiv preprint arXiv:2108.06507*.

Happ, C. & Greven, S. 2018. Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association* 113(522): 649-659.

Haslett, J. 1992. Spatial data analysis - challenges. *Journal of the Royal Statistical Society Series D: The Statistician* 41(3): 271-284.

Hubert, M., Rousseeuw, P.J. & Segaert, P. 2015. Multivariate functional outlier detection. *Statistical Methods & Applications* 24(2): 177-202.

Ieva, F. & Paganoni, A.M. 2013. Depth measures for multivariate functional data. *Communications in Statistics-Theory and Methods* 42(7): 1265-1276.

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: 321-338.

Mateu, J. & Giraldo, R. 2021. *Geostatistical Functional Data Analysis*. New York: John Wiley & Sons.

Ojo, O., Fernández Anta, A. & Lillo, R.E. 2019. Improvements to the Massive Unsupervised Outlier Detection (MUOD) algorithm. In *III International Workshop on Advances in Functional Data Analysis*.

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.

Sun, Y. & Genton, M.G. 2011. Functional boxplots. *Journal of Computational and Graphical Statistics* 20(2): 316-334.


*Corresponding author; email: rossita@um.edu.my