

Outlier Detection of Air Quality for Two Indian Urban Cities Using Functional Data Analysis

Mohammad Ahmad*, Weihu Cheng, Zhao Xu, Abdul Kalam

Faculty of Science, Beijing University of Technology, Beijing, China Email: *mahmad.or@emails.bjut.edu.cn, chengweihu@bjut.edu.cn, zhaox@bjut.edu.cn, faisal.stats@gmail.com

How to cite this paper: Ahmad, M., Cheng, W.H., Xu, Z. and Kalam, A. (2023) Outlier Detection of Air Quality for Two Indian Urban Cities Using Functional Data Analysis. *Open Journal of Air Pollution*, **12**, 79-91.

https://doi.org/10.4236/ojap.2023.123005

Received: July 27, 2023 **Accepted:** August 26, 2023 **Published:** August 29, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

http://creativecommons.org/licenses/by/4.0/

Abstract

Human living would be impossible without air quality. Consistent advancements in practically every aspect of contemporary human life have harmed air quality. Everyday industrial, transportation, and home activities turn up dangerous contaminants in our surroundings. This study investigated two years' worth of air quality and outlier detection data from two Indian cities. Studies on air pollution have used numerous types of methodologies, with various gases being seen as a vector whose components include gas concentration values for each observation per-formed. We use curves to represent the monthly average of daily gas emissions in our technique. The approach, which is based on functional depth, was used to find outliers in the city of Delhi and Kolkata's gas emissions, and the outcomes were compared to those from the traditional method. In the evaluation and comparison of these models' performances, the functional approach model studied well.

Keywords

Functional Data Analysis, Outliers, Air Quality, Gas Emission, Classical Statistics

1. Introduction

Modern lifestyles invariably include energy use and its consequences. Anthropogenic causes of air pollution include the combustion of straw, coal, and gasoline, as well as emissions from industries, vehicles, airplanes and cans of aerosols. On a daily basis, hazardous gases such as CO, CO₂, Particulate Matter (PM), NO₂, SO₂, O₃, NH₃, and Pb are released into our environment. Chemical compounds and particles in air pollution have an influence on the physical wellbeing of animals, humans, and ecosystems. Pneumonia, lung cancer, coronary artery disease, and influenza are among some of the serious ailments that may be contracted by people as a result of exposure to air pollution. Poor air quality is the root cause of smog, aerosol formation, reduced eyesight, rising temperatures, acid rain, early death, and other contemporary environmental issues. Machine learning was implemented to forecast air pollution in Indian cities [1] [2]. The monitoring was carried out at eight areas covering rural, semi-urban, and metropolitan environments. Throughout the campaign, the semi-urban location (Sirsa) had the greatest average concentration of specific concerns. According to this multi-city study, semi-urban areas have the worst air quality during rabi crop residue burning and require special attention to solve air quality concerns in the Indo-Gangetic plain region [3]. Delhi's indoor environment is deteriorating day by day as a result of vehicle pollution, and the effect of vehicular frequency on Delhi's air quality remains unknown [4]. Ahmad M. et al. [5] conducted the outliers study of Yamuna River using SPC and classical statistics. Haque M. S. and Singh R.B. [6] performed a case study of human health owing to air pollution in Kolkata, India. Only 39.3% of respondents reported that their health was negatively impacted by outdoor (air) pollution, despite the fact that the pollution level has been classified as critical. Using function data analysis and the impact of the COVID-19 outbreak on air quality, the outlier identification case research was conducted in Gijon, Spain [7]. The area of statistics being examined is functional data analysis (FDA), and this research uses its techniques to model this kind of data. It was chosen to address the inefficiency of the traditional approaches for finding outliers in vectorial data, which was discovered by comparing the findings from box plots and statistical control charts, both of which were used in this research to test for the presence of outliers. Environmental engineering [8] [9] [10] [11] [12], industrial processes [13] [14], sensors [15] [16], and medical research [17] are just a few of the many sectors in which FDA is used today. The advantage of functional analysis is that it allows you to investigate the identification issue from a time-related viewpoint. This is done by turning a set of discrete measurements that change over time into mathematical functions. Additionally, the best method in this investigation was determined to be the functional outlier indicator offered by Dai et al. [18] according to outlyingness. This technique can detect outliers with greater precision and robustness because it uses two variables: mean directional outlyingness, which examines a curve's shift to the remainder, and variation of directed outlyingness, which examines a curve's form to the rest. Despite there are various approaches for detecting outliers, that is the Grubbs test [19] and the Jäntschi test [20], these generally contain a vectorial set. In this instance, both conventional and functional methodologies were applied, resulting in an examination and contrast of both data to identify the most effective course of action. Fewer researchers have expressed an interest in finding outliers in Indian air quality. The goal of this study was to detect outliers in Particular Matters (PM2.5 and PM10) and trace gases (NO₂, NO_x, SO₂, and O₃) in Delhi and Kolkata, India. Kolkata is located in eastern India, while Delhi is located in northern India. The functional approach of data analysis was effectively used in this study. The dataset was treated as continuous measurements, and measurements over time without discrete values, using this novel functional technique. In order to assess the trend and periodicity of the data, this approach offers functional outliers that do not consider those discrete measurement errors to be outliers. Additionally, the functional method has the benefit of not requiring a data set that is normally distributed in comparison to most criteria for outliers' detection based on a discrete approach. As a result, further processing is not required. In this study, functional data analysis was used and compared to the traditional technique. This research is divided into sections, the second of which explains the material and techniques. Section 3 presents the methodology's results and, lastly, the key conclusion made from the results.

2. Materials and Methods

2.1. Materials

Data

The current investigation is being done to find outliers in two Indian cities, Delhi and Kolkata. Delhi is located in northern central India and has a population of 32.94 million people. Its size is 1483 km². This places it well ahead of Kolkata on the Asian continent. As a result, because Delhi lacks aquatic bodies that border it, temperature regulation occurs infrequently. Due to the lack of temperature regulation by the marine bodies, Delhi endures high temperatures both during the summer and winter seasons. Temperatures in Delhi often fluctuate between 2 and 45 degrees Celsius. Overall, Delhi's climate is a combination of semi-arid conditions mixed with monsoon-influenced humid subtropical conditions, with notable changes in summer and winter temperatures as well as variability in precipitation. On the other hand, Kolkata, which is in eastern India and has a populace of 15.33 million people and an area of 1886.67 km², is somewhat nearer to the sea and hence receives temperature regulation from the water bodies. Kolkata hence has more moderate temperature changes, and the transition from summer to winter there is controlled. Delhi is located at the other extreme of the temperature range.

Data for this study was provided by the air quality station. Daily station records over a period of two years, starting on January 1, 2018, and ending on December 31, 2019, make up the investigated data. As a consequence, there are 720 days since every month has been linearly extrapolated to have 30 days in order to meet the criteria for converting discrete data to functional data. Both cities' PM2.5, PM10, NO₂, NOx, O₃, and SO₂ levels were examined. μ g/m³ units are used to measure each variable. The data was obtained from India's Central Pollution Control Board (CPCB).

2.2. Methods

2.2.1. Classical Statistics

The objective of the conventional quantitative analysis is to examine the tendency, decide if any of them exceeds the limit, and track changes in air quality using descriptive statistics, such as mean, quartiles (Q1, Q2, Q3), time series, box plots, etc. A boxplot, a standardized technique that utilizes a five-number summary ("minimum, "Q1", median, "Q3", and "maximum"), is used to display the distribution of data. This can draw conclusions regarding the values of the outliers. We may discover that certain distributions or data sets require more details than the measures of central tendency. We need to understand the fluctuation and dispersion of the data. A boxplot is an illustration that successfully depicts the distribution of the values contained in the data.

2.2.2. Functional Data Analysis

FDA is a collection of methods for studying curves and functions to analyze data across time [21]. Begin by converting vector samples into functional samples. The beginning points, which come from the study's generated discrete values, are used to create the curves. Smoothing is the process of transforming vector points into a continuous function over time. This data composition is valuable in the research of air pollution since it takes all of the values from the day as a single unit. As a result, a day with NO₂, NOx, SO₂, O₃, PM2.5, and PM10 values of varying variability may have an average identical to the other days, and the vectorial approach detect the outliers. These days would be identified as possible outliers by the functional analysis. For outlier detection in these types of investigations, functional techniques have always been shown to be superior.

Let $x(t_f)$ represent the initial observations, $t_f \in R$ signifies the time steps, and p represents the number of observations ($f = 1, 2, \dots, p$). The individual value of the function $x(t) \in x \subset F$, where F is a functional space, can be observed. The functional space $F = span(\phi_1, \phi_2, \dots, \phi_p)$ is used to estimate x(t), where ϕ_g is the set of basis functions ($g = 1, 2, \dots, n_b$) and p is the number of basis functions necessary to generate a functional sample. In statistics, there are various types of bases, but the Fourier basis is the most commonly employed. Furthermore, for periodic data like the ones we have in our study, the Fourier basis is the best option [22].

$$\min_{x \in F} \sum_{f=1}^{p} \left(z_f - x \left(t_f \right) \right)^2 + \lambda \Gamma(x)$$
(1)

 $z_f = x(t_f) + \epsilon_f$ Where x is the observing point at t_f , ϵ_f is the random noise with zero mean, λ is the level of regulization and Γ is penalized, operator.

$$x(t) = \sum_{g=1}^{p} c_g \phi_g(t)$$
⁽²⁾

where $\{c_g\}_{g=1}^p$ is the coefficient that multiplied the basis function. We can write the problem of smoothing as

$$\min_{c} \left\{ \left(z - \phi_{c} \right)^{\mathrm{T}} \left(z - \phi_{c} \right) + \lambda c^{\mathrm{T}} R c \right\}$$
(3)

 $z = (z_1, \dots, z_p)^T$, the expansion of vector coefficient $c = (c_1, \dots, c_p)^T$, a (p, n_b) -matrix ϕ whose elements are $\phi_{fg} = \phi_g(t_f)$; and a (p, n_b) -matrix R whose elements are:

$$R_{gl} = \left\langle D^2 \phi_g, D^2 \phi_l \right\rangle_{L_2(T)} = \int_T D^2 \phi_g, D^2 \phi_l \mathrm{d}t$$
(4)

The problem can be solved with $c = (\phi'\phi + \lambda R)^{-1} \phi z$.

The functional data allows us to determine whether or not different time intervals, such as days, weeks, or months, are higher than the mean feature and how far they differ. It also enables the removal of outliers that aren't real but are caused by system failure. The notion of depth allows you to sort a collection of data in Euclidian space by how close it is to the sample core. In multivariate analysis, the concept of depth emerged and was generated to calculate a point centrality among a cloud of them. This idea started to be incorporated into practical data analysis over the course of the year. In this region, the centrality of a certain curve x_i is defined by depth, and the center of the sample is the mean curve. The two-depth measurement Fraiman-Muniz depth (FMD) and H-model depth (HMD) [22] are most usual in the sense of functional data.

Through the estimation of depths, it is also possible to classify outliers with a practical approach. In this case, it will take into account elements which have various behavioral designs than the rest. Instead of summarizing the curve observations into a single point, such as the average, the definition of depth makes it possible to deal with observations identified at a given interval in curve types. The depth technique uses for the identification of outlier and significance: there will be a low depth of an element that is distant from the sample. Thus, practical outliers are the curves with the least depth.

Firstly, the $F_{n,t}(x_e(t))$ is the cumulative empirical distribution function of the values of the curves $\{x_e(t)\}, (e=1,2,\dots,n)$ in a certain time $t \in [a,b]$ it is contemplated. It can be defined as:

$$F_{n,t}\left(x_{e}\left(t\right)\right) = \frac{1}{n} \sum_{g=1}^{n} I\left(x_{g}\left(t\right) \le x_{e}\left(t\right)\right)$$

$$\tag{5}$$

where I(.) is an indicator function, next, the FMD for curve x_i is calculated as:

$$\operatorname{FMD}_{n}\left(x_{e}\left(t\right)\right) = \int_{a}^{b} D_{n}\left(x_{e}\left(t\right)\right) \mathrm{d}t \tag{6}$$

where $t \in [a,b]$. The functional mode in HMD, on the other hand, is the element or curve that is most densely surrounding by the other curves in the dataset. HMD is written as:

$$\operatorname{HMD}_{n}(x_{e},h) = \sum_{g=1}^{n} K\left(\frac{\left\|x_{e} - x_{g}\right\|}{h}\right)$$
(7)

In a functional space, with a kernel function $K: R^+ \to R^+$, a bandwidth parameter *h* and $\|\cdot\|$ as the norm. In a vast majority of cases, it is norm L₂, expressed as:

$$\|x_{e}(t) - x_{f}(t)\| = \left(\int (x_{e}(t) - x_{f}(t))^{2} dt\right)^{1/2}$$
(8)

There are also a number of parameters for the kernel functions $K(\cdot)$. The truncated Gaussian kernel is a popular one, can be expressed as:

$$K(t) = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right), \ t > 0 \tag{9}$$

A functional sample set may have elements that, although not containing error, exhibit characteristics that are distinct from the rest of the set. Instead of only comparing the mean values over the measurement time interval, the depth measurements mentioned above allow sets of observations over time fitted to curves to be compared in order to find outliers in functional samples. An outlier in a functional sample will therefore have significantly less depth because depth and outlier are opposite terms. In order to find functional outliers, the deepest curves are sought after. The value of bandwidth h was chosen as the 15th percentile of the empirical distribution using the HMD to create the outlier selection criterion $\left\{ \left\| x_e(t) - x_f(t) \right\|^2, e, f = 1, 2, \dots, n \right\}$ [23]. The cut-off C was chosen so that around 1% of accurate observations were incorrectly classified as outliers (type I error) [24]:

$$P_r(\text{HMD}_n(x_e(t))) < c = 0.01, \ e = 1, 2, \cdots, n$$
 (10)

Unfortunately, the distribution of the selected functional depth is unknown, necessitating an estimate of C. For the purposes of this study, we selected a method based on bootstrapping [24] [25] [26] the curves of the original set with a probability proportional to depth out of the several approaches to estimate this value. As an overview, the bootstrapping strategy is as stated:

1) By using sampling with replacement, a new sample is taken from the previous sample (each element is replaced after extraction so it can be chosen again). In addition, order 10 has been chosen for resampling.

2) The populational parameter of interest is estimated using this new sample as a basis to generate a statistic.

- 3) Repeat the steps overhead a significant number of times.
- 4) Finally, determine the empirical statistical distribution.

3. Results

The process of obtaining and assessing the findings took place across two stages. In the first stage, box plots were used to visually examine each database variable for air quality. The construction of the functional chart with daily measurement preferred over monthly continued in the second phase. The daily information gathered from the CPCB for a 24-month period (from January 1 2018 to December 31, 2019). Hence the result, our research focuses on the study of daily NO₂, NOx, SO₂, O₃, PM2.5, and PM10 data in Delhi and Kolkata. The boxplot in **Figure 1(a)**, **Figure 1(b)** presents the data group of NO₂, NOx, SO₂, O₃, PM2.5, and PM10 concentrations by quartiles. The upper line of the box represents the third quartile (Q3), the second line is the median, the bottom line represents the first quartile (Q1), and the red dots represent the outlier for both Delhi and Kolkata.

The variable that we were selected in this study, our sample $\{x_{ij}\}_{i=1}^{24}$ j=1

corresponded to the 24 months (January 2018 to December 2019) for all variable, where x_{ij} is the emission measurement for day *i* of month *j*, $i = 1, 2, \dots, 30$. Following the smoothing mentioned above, a sample $\{x_i\}_{i=1}^{2^4}$ is created, where each x_i is now a function, taking into account a set of 1000 basis elements. This procedure yields a correlation score of 99% for the discrete values and the measured values of each function at each location. In this regard, the produced functional sample has a 99% correlation with the discrete sample. Figure 2 (left side) shows that the functional representation of the NO₂, NO_x, O₃, SO₂, PM2.5, and PM10 and functional outlier detection Figure 2 (right side). There are numerous outliers obtained using functional approach and is shown in Table 1, this table shows that Delhi has more outlier than Kolkata. These curves were identified as outliers due to weather circumstances. Outliers are detected as a result of several causes such as traffic pollution, gasoline consumption, firework exploding during Diwali, stubble burning, and road dust. In certain conditions, the conventional technique discovers a huge number of outliers, even finding outliers in more than 50% of the data. This issue occurs regardless of whether the station or compound is being reviewed. The number of outliers detected by functional analysis is substantially fewer in both Delhi and Kolkata (NO₂, NOx, O₃, SO₂, PM2.5, and PM10), as shown in the figure that precedes functional outlier identification.



Figure 1. Box plot representation of Delhi (a) and Kolkata (b) NO₂, NO_x, SO₂, O₃, PM2.5, and PM10.

Variable	Outliers (Delhi)	Outliers (Kolkata)
NO ₂	3	0
NOx	7	2
O ₃	3	0
SO ₂	1	2
PM2.5	4	2
PM10	4	0

 Table 1. Functional Data Analysis outlier results from the collected dataset (Delhi and Kolkata).







Figure 2. Functional representation of data (left) and functional outlier (right) [a-x] for Delhi and Kolkata NO₂, NOx, SO₂, O₃, PM2.5, and PM10. The black dot curve represents the outlier.

4. Conclusion

Due to the environment's ongoing changes and unpredictable nature, and because contaminants vary in location and time, forecasting air quality is a tough task. Continuous air quality monitoring and research are required, particularly in developing countries, due to the harmful effects of air pollution on people, animals, plants, historical sites, the climate, and the environment. Nevertheless, researchers have shown little interest in AQI prediction for India. A traditional approach and a functional technique that consider data as functions are employed in this study endeavour. Compared to the vector technique that examines means and cannot take into account temporal fluctuations, the functional strategy has the benefit of allowing more details to be obtained from the data. Functional Data Analysis, which offers significant benefits over conventional methods in the detection of particular variability if data differ from a normal distribution, may be used to discover outliers in non-normal data sets. To assess the quality of the air in metropolitan areas, outlier identification using functional data analysis might be utilized. Pollutant emissions often rise along with economic and population expansion and fall during economic contractions. In this study, a novel approach for detecting outliers in the collection of gas emissions was devised. While other cities with comparable or dissimilar sources of pollution can use these strategies, it is always important to consider the unique characteristics of each place.

Funding

This work was supported by the National Natural Science Foundation of China (Grant Number 11801019) and Beijing Natural Science Foundation (Grant Number Z190021).

Data Availability Statement

The data has been made publicly available by the Central Pollution Control Board: <u>https://cpcb.nic.in/</u> which is the official portal of Government of India. They also have a real-time monitoring app: <u>https://app.cpcbccr.com/AQI India/</u>.

Acknowledgements

The authors would like to thank the editor as well as the reviewers for their many insightful and helpful comments and suggestions that greatly helped the paper. I would like to convey my heartfelt gratitude to Weihu Cheng for his insightful and helpful advice during the inception and implementation of this research project. The excellent efforts and support of Mohammad Nuruzzama and Zhao Xu were likewise greatly appreciated.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- Kumar, K. and Pande, B.P. (2023) Air Pollution Prediction with Machine Learning: A Case Study of Indian Cities. *International Journal of Environmental Science and Technology*, 20, 5333-5348. <u>https://doi.org/10.1007/s13762-022-04241-5</u>
- Persis, J. and Amar, A.B. (2023) Predictive Modeling and Analysis of Air Quality—Visualizing before and during COVID-19 Scenarios. *Journal of Environmental Management*, **327**, Article ID: 116911. https://doi.org/10.1016/i.jenyman.2022.116911
- [3] Ravindra, K., Singh, T., Singh, V., Chintalapati, S., Beig, G. and Mor, S. (2023) Understanding the Influence of Summer Biomass Burning on Air Quality in North India: Eight Cities Field Campaign Study. *Science of the Total Environment*, 861, Article ID: 160361. <u>https://doi.org/10.1016/j.scitotenv.2022.160361</u>
- [4] Kumar, V., Gupta, S. and Jolli, V. (2022) Influence of Vehicular Frequency on Air Quality of Delhi, India. *Ecological Chemistry and Engineering S*, 29, 477-485. <u>https://doi.org/10.2478/eces-2022-0034</u>
- [5] Ahmad, M., Haq, A., Kalam, A. and Shah, S.K. (2022) A Comparative Study of Outlier Detection of Yamuna River Delhi India by Classical Statistics and Statistical Quality Control. *Reliability: Theory & Applications*, 17, 430-438.
- [6] Haque, M.S. and Singh, R.B. (2017) Air Pollution and Human Health in Kolkata, India: A Case Study. *Climate*, 5, Article No. 77. <u>https://doi.org/10.3390/cli5040077</u>
- [7] Rigueira, X., Araújo, M., Martínez, J., García-Nieto, P.J. and Ocarranza, I. (2022) Functional Data Analysis for the Detection of Outliers and Study of the Effects of the COVID-19 Pandemic on Air Quality: A Case Study in Gijón, Spain. *Mathematics*, **10**, Article No. 2374. <u>https://doi.org/10.3390/math10142374</u>
- [8] Febrero, M., Galeano, P. and González-Manteiga, W. (2008) Outlier Detection in Functional Data by Depth Measures, with Application to Identify Abnormal NO_x Levels. *Environmetrics*, 19, 331-345. <u>https://doi.org/10.1002/env.878</u>
- [9] Matías, J.M., Ordóñez, C., Taboada, J. and Rivas, T. (2009) Functional Support Vector Machines and Generalized Linear Models for Glacier Geomorphology Analysis. *International Journal of Computer Mathematics*, 86, 275-285. https://doi.org/10.1080/00207160801965305
- [10] Torres, J.M., Nieto, P.G., Alejano, L. and Reyes, A.N. (2011) Detection of Outliers in Gas Emissions from Urban Areas Using Functional Data Analysis. *Journal of Hazardous Materials*, 186, 144-149. <u>https://doi.org/10.1016/j.jhazmat.2010.10.091</u>
- [11] Martínez, J., Saavedra, Á., García-Nieto, P.J., Piñeiro, J.I., Iglesias, C., Taboada, J., Sancho, J. and Pastor, J. (2014) Air Quality Parameters Outliers Detection Using Functional Data Analysis in the Langreo Urban Area (Northern Spain). *Applied Mathematics and Computation*, 241, 1-10. https://doi.org/10.1016/j.amc.2014.05.004
- [12] Sancho, J., Iglesias, C., Piñeiro, J., Martínez, J., Pastor, J.J., Araújo, M. and Taboada, J. (2016) Study of Water Quality in a Spanish River Based on Statistical Process Control and Functional Data Analysis. *Mathematical Geosciences*, 48, 163-186. <u>https://doi.org/10.1007/s11004-015-9605-y</u>
- [13] Ordòñez, C., Martìnez, J., Saavedra, À. and Mourelle, A. (2011) Intercomparison Exercise for Gases Emitted by a Cement Industry in Spain: A Functional Data Approach. *Journal of the Air & Waste Management Association*, **61**, 135-141. <u>https://doi.org/10.3155/1047-3289.61.2.135</u>
- [14] Sancho, J., Pastor, J.J., Martínez, J. and García, M.A. (2013) Evaluation of Harmonic

Variability in Electrical Power Systems through Statistical Control of Quality and Functional Data Analysis. *Procedia Engineering*, **63**, 295-302. https://doi.org/10.1016/j.proeng.2013.08.224

- [15] Wu, D., Huang, S. and Xin, J. (2008) Dynamic Compensation for an Infrared Thermometer Sensor Using Least-Squares Support Vector Regression (LSSVR) Based Functional Link Artificial Neural Networks (FLANN). *Measurement Science and Technology*, **19**, Article ID: 105202. https://doi.org/10.1088/0957-0233/19/10/105202
- [16] Galán, C.O., Torres, J.M., de Cos, F.J. and Lasheras, F.S. (2011) Comparison of GPS Observations Made in a Forestry Setting Using Functional Data Analysis—CMMSE 2010. *International Journal of Computer Mathematics*, **89**, 402-408.
- [17] Dombeck, D.A., Graziano, M.S. and Tank, D.W. (2009) Functional Clustering of Neurons in Motor Cortex Determined by Cellular Resolution Imaging in Awake Behaving Mice. *Journal of Neuroscience*, **29**, 13751-1360. https://doi.org/10.1523/JNEUROSCI.2985-09.2009
- [18] Dai, W. and Genton, M.G. (2018) Multivariate Functional Data Visualization and Outlier Detection. *Journal of Computational and Graphical Statistics*, 27, 923-934. <u>https://doi.org/10.1080/10618600.2018.1473781</u>
- [19] Grubbs, F.E. (1969) Procedures for Detecting Outlying Observations in Samples. *Technometrics*, 11, 1-21. <u>https://doi.org/10.1080/00401706.1969.10490657</u>
- [20] Jäntschi, L. (2019) A Test Detecting the Outliers for Continuous Distributions Based on the Cumulative Distribution Function of the Data Being Tested. *Symmetry*, **11**, Article No. 835. <u>https://doi.org/10.3390/sym11060835</u>
- [21] Ramsay, J. and Silverman, B. (2005) Functional Data Analysis. Springer, New York. https://doi.org/10.1007/b98888
- [22] Martínez Torres, J., Pastor Pérez, J., Sancho Val, J., McNabola, A., Martínez Comesaña, M. and Gallagher, J. (2020) A Functional Data Analysis Approach for the Detection of Air Pollution Episodes and Outliers: A Case Study in Dublin, Ireland. *Mathematics*, 8, Article No. 225. <u>https://doi.org/10.3390/math8020225</u>
- [23] Cuevas, A. and Fraiman, R. (1997) A Plug-In Approach to Support Estimation. *The Annals of Statistics*, 25, 2300-2312. <u>https://doi.org/10.1214/aos/1030741073</u>
- [24] Cuevas, A., Febrero, M. and Fraiman, R. (2006) On the Use of the Bootstrap for Estimating Functions with Functional Data. *Computational Statistics & Data Analysis*, 51, 1063-1074. <u>https://doi.org/10.1016/j.csda.2005.10.012</u>
- [25] Febrero, M., Galeano, P. and González-Manteiga, W. (2007) A Functional Analysis of NO_x Levels: Location and Scale Estimation and Outlier Detection. *Computational Statistics*, 22, 411-427. <u>https://doi.org/10.1007/s00180-007-0048-x</u>
- [26] Peng, L. and Qi, Y. (2008) Bootstrap Approximation of Tail Dependence Function. Journal of Multivariate Analysis, 99, 1807-1824. https://doi.org/10.1016/j.jmva.2008.01.018