mathematics

MDPI

Article

Method for the Detection of Functional Outliers Applied to Quality Monitoring Samples in the Vicinity of El Musel Seaport in the Metropolitan Area of Gijón (Northern Spain)

Luis Alfonso Menéndez-García [1], Paulino José García-Nieto [2], Esperanza García-Gonzalo [2], Fernando Sánchez Lasheras [2], Laura Álvarez-de-Prado [1,*] and Antonio Bernardo-Sánchez [1]

[1] Department of Mining Technology, Topography and Structures, Higher and Technical School of Mining Engineering, University of León, Campus de Vegazana s/n, 24071 León, Spain; lmeneg00@estudiantes.unileon.es (L.A.M.-G.); abers@unileon.es (A.B.-S.)

[2] Department of Mathematics, Faculty of Sciences, University of Oviedo, 33007 Oviedo, Spain; pjgarcia@uniovi.es (P.J.G.-N.); espe@uniovi.es (E.G.-G.); sanchezfernando@uniovi.es (F.S.L.)

* Correspondence: laura.alvarez@unileon.es; Tel.: +34-987-291-959

Abstract: Air pollution affects human health and is one of the main problems in the world, including in coastal cities with industrial seaports. In this sense, the city of Gijón (northern Spain) stands out as one of the 20 Spanish cities with the worst air quality. The study aims to identify outliers in air quality observations near the El Musel seaport, resulting from the emissions of six pollutants over an eight-year period (2014–2021). It compares methods based on the functional data analysis (FDA) approach and vector methods to determine the optimal approach for detecting outliers and supporting air quality control. Our approach involves analyzing air pollutant observations as a set of curves rather than vectors. Therefore, in the FDA approach, curves are constructed to provide the best fit to isolated data points, resulting in a collection of continuous functions. These functions capture the behavior of the data in a continuous domain. Two FDA approach methodologies were used here: the functional bagplot and the high-density region (HDR) boxplot. Finally, outlier detection using the FDA approach was found to be more powerful than the vector methods and the functional bagplot method detected more outliers than the HDR boxplot.

Keywords: outlier detection; functional bagplot; functional high-density region (HDR) boxplot; air pollution

MSC: 62R10; 62P12; 62G32

Citation: Menéndez-García, L.A.; García-Nieto, P.J.; García-Gonzalo, E.; Lasheras, F.S.; Álvarez-de-Prado, L.; Bernardo-Sánchez, A. Method for the Detection of Functional Outliers Applied to Quality Monitoring Samples in the Vicinity of El Musel Seaport in the Metropolitan Area of Gijón (Northern Spain). *Mathematics* **2023**, *11*, 2631. https://doi.org/10.3390/math11122631

Academic Editor: Jose Luis Vicente Villardon

Received: 4 April 2023
Revised: 4 June 2023
Accepted: 6 June 2023
Published: 8 June 2023

1. Introduction

Air pollution is a global problem with significant health implications [1]. The growth in maritime transport due to globalization has made seaports crucial for economic and social development, accounting for over 80% of world trade [2,3]. Several ports in northwest Spain, including La Coruña, Ferrol, and Gijón (El Musel), have expanded to accommodate increased traffic [4,5]. Pollution issues have also been observed in other Spanish ports such as Santander and La Coruña, which have implemented various solutions to mitigate the problem [6].

Port activities, including ship maneuvers, fueling, and docking, contribute to substantial pollution, especially during non-optimal engine operation phases [7,8]. Ships, particularly those with diesel engines, emit significant amounts of particulate matter, both primary and secondary, which can have a substantial impact on nearby populations [9,10]. Industrial ports additionally contribute to air pollutant emissions through discharges, handling of bulk materials, road dust, and heavy vehicle traffic, all influenced by meteorological conditions, material types, and proximity to populations [11,12]. In some cases,

traffic within the port area can have a more significant impact on air quality than bulk unloading [13].

All these activities have been identified as sources of air pollution that can negatively affect human health, leading to the exceeding of daily limit values for inhalable particles [14,15]. Numerous studies have examined the impact of air quality in cities with seaports and the transport of pollution inland [16,17]. Coastal areas, especially those downwind of port areas, experience specific meteorological patterns that affect the dispersion, transformation, accumulation, and deposition of air pollutants [18].

Several studies have evaluated the impact of loading and unloading operations on particulate matter emissions in ports, such as those conducted in the ports of Tarragona and Thessaloniki [12,19]. Other studies have examined PM_{10} (particles smaller than 10 microns) and $PM_{2.5}$ (particles smaller than 2.5 microns) emissions resulting from port activities in Barcelona and Civitavecchia, as well as noise and pollution caused by ship and port activities in Ancora [20–22]. Pollution from ships in ports such as those in Shanghai and Istanbul has been analyzed, along with studies on pollutant gas compositions, PM_{10} levels due to bulk handling in Alicante, and the relationship between maritime traffic and pollutant concentration in several ports [17,23].

It is well-known that there are different types of outliers, such as vector and functional outliers. To date, there have been a number of studies on the identification of vector outliers in fields as diverse as medicine [24] and renewable energy [25]. Numerous methods have been proposed to detect and investigate outliers in pollutant measurements recorded at air quality monitoring stations [26–28]. However, no single method has been universally accepted as the definitive approach.

Outlier detection based on vectorial methods, such as Z-score and Rosner test, involves analysis of the distribution of individual data points with respect to the mean and standard deviation of the data set. On the other hand, methods based on the FDA approach are designed to analyze functional data sets, where each observation is a curve or function. These outlier detection methods are Bagplot and HDR Boxplot.

In the area surrounding the port of Gijón, there have been recent and repeated episodes of pollution that have caused social alarm and concern about possible health effects. Municipal protocols to combat pollution have also been activated on several occasions. The objective of this study was to propose a model for the identification of outliers derived from pollutant emissions into the atmosphere in the vicinity of the industrial port of El Musel in Gijón (latitude 43°34′ N, longitude 5°41′ W) using different methodologies. These included vectorial methods and methods based on the FDA approach, allowing a comparative analysis between the two approaches. In this way, early warning indicators can be provided to help local authorities make timely decisions.

The structure of this paper is as follows: Section 2 introduces the experimental setup and variables involved in this study. Section 2.4 explains the mathematical methods used in this paper, including Box–Cox transformation, outlier detection using vector methods, smoothing technique, functional bagplot, and HDR boxplot. Section 3 describes the results obtained with the proposed models and provides a discussion by comparing both methodologies. Finally, Section 4 concludes this study by highlighting the main findings.

2. Materials and Methods

2.1. Gijón and the Port of El Musel

Gijón, with a population of 270,871 registered in 2021, is located on the coast of the Cantabrian Sea (Figure 1a), in the autonomous community of Principality of Asturias, in northern Spain. Gijón's climate is characterized by short, mild summers with an average temperature of 18.7 degrees Celsius (°C) in August. Winters are long, wet, and relatively cold, with an average minimum temperature of 8.1 °C in January and February, which are the coldest months. It rains throughout the year in Gijón, with July being the driest month with an average of 72 millimeters (mm) and November being the wettest with an average of 159 mm. The total average annual rainfall is 1303 mm and the average annual

temperature is 13.0 °C. Prevailing wind conditions are SW in winter and NE in summer. The strongest wind conditions are NW.

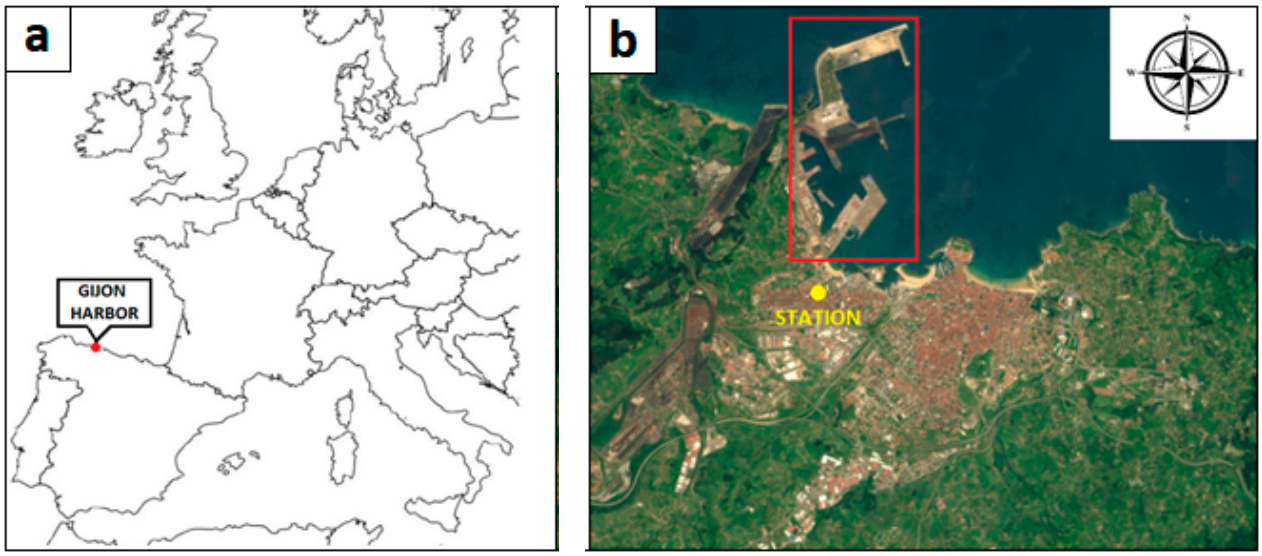

Figure 1. (**a**) Position of the port of Gijón; and (**b**) location of the Argentina air quality monitoring station on map.

The construction of the Port of El Musel in Gijón was due to the impulse of the iron and steel industry and the exploitation of hard coal in the second quarter of the 19th century in Asturias. Commercial exploitation began in 1907 [29]. It has undergone numerous expansions throughout its history, the most recent of which took place between 2005 and 2011. Currently, El Musel is an artificial outer harbor separated from the urban area of the city of Gijón, whose main economic activity is the traffic of goods, connected to more than 200 ports worldwide. The port is composed of seven main piers. Figure 1b shows its location in relation to the city. The total length of quays and berths in the commercial basins are 6934 meters (m) and 1006 m in fishing basins. The main materials unloaded include coal, petroleum coke, clinker, iron ore, and other non-metallic materials. It is the largest bulk port in Spain, with 14,500 thousand tons (kt) in 2019 and 13,616 kt of solid bulk in 2020 [30,31]. This can cause a large amount of air pollution, which is further exacerbated by road traffic along the port facilities and the roads leading to and from the port.

2.2. Pollution

Air pollutants can be divided into two types. The first type is primary pollutants, which are emitted directly from sources into the atmosphere. The second type is secondary pollutants, which are formed from primary pollutants via chemical reactions and microphysical processes in the atmosphere [32].

The origin of pollutants can be attributed to human activities. In this case, they are called anthropogenic. Otherwise, they are referred to as naturally occurring pollutants.

Sulfur dioxide (SO_2) is produced, among other sources, by exhaust emissions from diesel engines, including those from ships, vehicles, and industrial plants [33]. Aerosols formed via reaction with water vapor are hazardous to health even at low concentrations and can cause irritation in the respiratory system [32]. They also cause damage to structures by contributing to corrosion problems through the formation of sulfuric acid and producing acid rain [34].

Carbon monoxide (CO) is a gas produced primarily by anthropogenic sources, such as incomplete combustion from road traffic [35].

Ozone (O_3) is a highly irritating pollutant to humans, is harmful to plants, and is commonly found in urban smog. It is of secondary origin, formed through the reaction of

NO_x (nitrogen oxides) and VOCs (volatile organic compounds) with sunlight. It causes numerous respiratory and health problems, as reported in numerous studies [36,37].

PM10 particles are small enough to pass through the respiratory tract and lodge in the lungs. Short-term exposure can aggravate existing health issues, while long-term exposure may lead to lung disease and other ailments [38].

The most harmful nitrogen oxides as pollutants are NO_2 and NO (nitrogen dioxide and nitric oxide) [32]. Extensive research has shown the harmful effects of these pollutants on human health, affecting both adults [39] and children [40]. They can react with water vapor to form nitric acid, which, like sulfur oxides, causes metal corrosion issues [41].

2.3. Experimental Data Set

In this article, the values of the pollutants PM_{10}, NO, NO_2, SO_2, CO, and O_3 collected over six years, from January 2016 to December 2021, at the Principality of Asturias monitoring station ES1271A, located in Gijón, have been analyzed. This air quality monitoring station, called Argentina, is the closest to the port, as shown in Figure 1b.

2.4. Mathematical Methodologies Used in This Study

2.4.1. Box–Cox Transformation

In order to carry out the study of vector outliers, a normality condition was required, which was not present in the initial data recorded by the air quality monitoring station of this study. One of the objectives of variable transformation is to reduce skewness as well as other distributional properties that cause difficulties in statistical analysis. Therefore, it was necessary to find a simple transformation that could approximate the data of a variable to a normal distribution. Linear transformations do not change the shape measures of skewness and kurtosis [42]. However, nonlinear transformations can modify these measures and can be used to obtain variables with distributions that approximate normal distribution. Among these transformations, the most commonly used is the Box–Cox transformation, which belongs to the following family of transformations [43]:

$$x_j^{(\lambda)} = \begin{cases} \frac{x_j^{\lambda}-1}{\lambda}, & if\ \lambda \neq 0, \\ \log(x_j), & if\ \lambda = 0 \end{cases} \quad (1)$$

where x_j represents the variable to be transformed and λ is the transformation coefficient that maximizes the likelihood function of the variable's profile. This makes it possible to perform a transformation that allows the data of a study to be treated as approximately normal in most cases [44].

Subsequently, after applying the Box–Cox transformation, normality was assessed using the Anderson–Darling test, a widely known test described in [45].

2.4.2. Outlier Detection Techniques Using Vector Methods

Z-Score

One methodology employed to identify atypical values (or outliers) is well-known as the Z-score, which assumes data normality [46]. This approach identifies outliers as standardized values that exceed a given threshold. The Z-score is calculated as follows:

$$z_i = \frac{x_i - x_r}{\sigma}, \quad (2)$$

where x_i represents the value of the observation at each point, x_r is a reference value (typically the mean of the observations), and σ is the standard deviation of all observations. The detailed development of this methodology can be found in [46].

The scoring criterion specified by the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC 17043:2010) was adopted as the criterion to establish possible outliers [47]:

$$\begin{array}{l} |z_i| < 2 \rightarrow \text{Satisfactory;} \\ 2 \leq |z_i| \leq 3 \rightarrow \text{Uncertain;} \\ |z_i| > 3 \rightarrow \text{Unsatisfactory.} \end{array} \quad (3)$$

Therefore, a Z-score greater than two was considered a possible outlier, while a score greater than three indicated a clear outlier. It is important to note that this criterion assumes a normal distribution of the contamination values, necessitating performing a prior transformation using the Box–Cox methodology.

A set of observations may have upper and lower outliers, which are, respectively, significantly higher or lower than the remaining values. In the case of air quality studies, higher concentrations of pollutants are of particular concern to human health, and therefore require monitoring. By applying the Z-score methodology, positive and negative values can be obtained. Following the criteria given in Equation (3) to identify possible outliers, we identified higher outliers ($z_i > 3$) and lower outliers ($z_i < 3$). However, for the purpose of this study, only positive values ($z_i > 3$) were considered. Since our focus was on air quality, our interest was only in the upper outliers.

Rosner's Test

Rosner's test is a statistical tool used to detect outliers within a data set. Its basic principle is to compare the difference between a given value and the mean of the remaining values, which is then divided by the standard deviation of those values. A larger standardized difference indicates a higher probability that the value is an outlier.

The test is performed iteratively, meaning that the procedure is repeated for each potential outlier value, with the value with the largest standardized difference being eliminated at each iteration.

Furthermore, the specific method employed in Rosner's test is known as the generalized ESD (Extreme Studentized Deviate) Test. This approach addresses certain limitations found in the Grubbs Test, also referred to as ESD or Pearson-Hartley. The ESD Test was developed to tackle two primary limitations: first, it was designed to test a single outlier and cannot find multiple outliers in a data set; second, it can suffer from a masking effect due to the presence of other outliers nearby [48].

Rosner's test is designated to detect one or more outliers in a data set having an approximately normal distribution. It requires specifying an upper limit, denoted as r, for the number of outliers to be tested [49]. The generalized ESD performs r separate tests, looking for from one to r outliers. This approach avoids the masking effect and allows for the detection of multiple outliers [49].

Rosner's test hypotheses are as follows [50]:

H_0: *There are no outliers in the given data set, i.e., the data set is normally distributed.*

H_1: *There are up to a maximum of r outliers in the given data set.*

The development of Rosner's test, indicated in [49,51], is as follows:

Let $x_1, x_2, \ldots, x_n$ represent the n observations of a given data set X. It is considered that $n - r$ of the observations of that data set come from a normal distribution, while the r observations are potentially extreme and may or may not belong to a different distribution.

Let us now consider a new data set X^* consisting of the elements $x_1^*, x_2^*, \ldots, x_{n-i}^*$, where $n - i$ are the number of observations after having removed the i most extreme measurements from the previous set X; therefore, let $i = 0, 1, 2, \ldots, r - 1$.

We denote the standard deviation as $s^{(i)}$ and the mean as $\bar{x}^{(i)}$ of the set X^*, as well as $\bar{x}^{(0)}$ and $s^{(0)}$ to the mean and standard deviation of the set X containing all observations.

We can define the most extreme observation in the data set, a specific value of i, as that which is furthest from the mean. That is:

$$x^{(i)} = max_j |x_j^* - \bar{x}^{(i)}|, j = 1, 2, \ldots, n-1. \quad (4)$$

Although, from the point of view of contaminant concentration analysis, the largest extreme observations are of interest, formally, from a mathematical point of view, the absolute value is taken to consider as an extreme observation the smallest or the largest of the elements of the data set.

Rosner's outlier test [49,50] is supported by r statistics, denoted as $R_1, R_2, \ldots, R_r$; $r \in \mathbb{N}/r < n$, corresponding to the studentized extreme deviations obtained from subsets of decreasing size, i.e., $n, n-1, \ldots, n-r+1$ elements.

Therefore, we can define the statistic R_{i+1} and the critical value λ_{i+1} as follows:

$$R_{i+1} = \frac{\left| x^{(i)} - \bar{x}^{(i)} \right|}{s^{(i)}}, \quad (5)$$

$$\lambda_{i+1} = \frac{t_{p,n-i-2}(n-i-1)}{\sqrt{(n-i-2+t_{p,n-i-2})(n-i)}}, \quad (6)$$

where $p = 1 - \frac{\alpha}{2(n-i)}$, $t_{p,\nu}$ is the p^{th} quantile with ν degrees of freedom (dof) of the t-Student distribution, and α is the Type I error level.

The procedure to establish the number of outliers is as follows: first, compare the statistic R_r with the critical value λ_r. If $R_r > \lambda_r$, then we can conclude that the most extreme r value is an outlier. If $R_r \leq \lambda_r$, proceed to compare R_{r-1} with λ_{r-1}. If $R_{r-1} > \lambda_{r-1}$, then the $r-1$ most extreme values are considered outliers. Finally, we continue in the same way until we have detected up to a certain number r of outliers or until the test does not give any value [50].

Rosner's test is indicated for sets with $n > 25$ observations [50].

2.4.3. Outlier Detection Methods Based on the FDA Approach

Functional data analysis (FDA) is a statistical framework designed for analyzing data in the form of smooth curves or functions. The basic philosophy of FDA is that, despite their discrete nature, observed data functions should be considered as single entities, rather than as a mere sequence of individual observations. A functional data set, denoted by x, consists of n pairs (t_j, Y_j), where Y_j represents a recording or observation of the function $x(t_j)$ at argument value t_j. Since observation noise is typically present in data, the representation of raw data in a functional form often involves a smoothing process. The FDA approach is developed in detail in Ramsay and Silverman [52].

Outlier detection within the FDA approach focuses on identifying unusual observations or curves that deviate significantly from the expected behavior. These outliers provide valuable insight into abnormal or extreme patterns and are of great interest in various fields. The goals of outlier detection using methods based on the FDA approach include identifying outliers of magnitude and shape, characterizing their impact on the data, and understanding the underlying mechanisms that contribute to their occurrence.

The following subsections describe smoothing in more detail, as well as the two outlier detection methods based on the FDA approach used in this study: functional bagplot and functional high-density region boxplot.

Smoothing: A Method to Represent Curves Based on Points

In certain problem domains, measurements are derived from operations that have functional characteristics. In the present study, the data under investigation were derived from the average daily concentrations of atmospheric pollutants, aggregated on a monthly basis, as recorded by an air quality monitoring station. The focus was on the analysis of a set of air pollutants observed over a consecutive period of months. Initial exploratory analysis revealed a discernible pattern of variation with a sinusoidal profile. This observed variation could potentially be influenced by meteorological factors such as temperature, rainfall, humidity, wind direction, and wind speed, as well as various processes associated with nearby industrial and port activities. The FDA approach represents an appropriate method for investigating these types of issues.

In the FDA approach, curves are created to fit the isolated data points, transforming a vector sample into a functional sample. This results in a set of continuous functions instead of multiple point values.

Applying appropriate smoothing assumptions in data analysis allows for a greater amount of information to be obtained and contributes to flexible and robust results. The smoothing approach must accurately represent the desired characteristics.

Smoothing and curve fitting are different concepts. Curve fitting aims to find a curve that closely resembles the original data.

According to Ramsay and Silverman [52], monitoring a continuous process at isolated points leads to monitoring functional data. Let F be a functional space and $x(t_j)$ be a set of n_p observations, where t_j represents each instant of time. These observations can be considered as the values taken by the function $x(t) \in \chi \subset F$ at specific points. For evaluation purposes, we assume $F = span\{\phi_1, \ldots, \phi_{n_b}\}$, where $\{\phi_k\}$ is a set of basis functions, $k = 1, 2, \ldots, n_b$, and n_b is the number of functions needed to form the basis of the functional space. Typically, spline or Fourier functions are used as basis functions. Although other types are possible [53], we considered a Fourier function:

$$x(t) = \sum_{k=1}^{n_b} c_k \phi_k(t), \tag{7}$$

where $\{c_k\}_{k=1}^{n_b}$ are the function coefficients of $x(t)$ with respect to a collection of selected basis functions. The smoothing problem is based on finding the solution to the following regularization issue [53]:

$$\min_{x \in F} \sum_{j=1}^{n_p} \{z_j - x(t_j)\}^2 + \lambda \Gamma(x). \tag{8}$$

In this expression, $z_j = x(t_j) + \varepsilon_j$ calculates the value of the function x at each point t_j, where ε_j is a random perturbation with zero mean. Γ is an operator that penalizes problem difficulty and λ is a regularization coefficient controlling the intensity.

A Fourier series can approximate periodic functions, both univariate and multivariate. Thus, the expansion in Expression (8) can be written as:

$$\min_{c} \Big\{ (\boldsymbol{z} - \boldsymbol{\Phi c})^T (\boldsymbol{z} - \boldsymbol{\Phi c}) + \lambda \boldsymbol{c}^T \boldsymbol{R c} \Big\}, \tag{9}$$

where $\boldsymbol{z} = \left(z_1 \ldots, z_{n_p}\right)^T$ is the vector from measurements, $\boldsymbol{c} = \left(c_1 \ldots, c_{n_p}\right)^T$ is the vector of the coefficients from the functional expansion, $\boldsymbol{\Phi}$ is an $n_p \times n_p$ matrix with elements $\boldsymbol{\Phi}_{jk} = \phi_k(t_j)$, and $\boldsymbol{R}$ is an $n_b \times n_b$ matrix defined as:

$$R_{kl} = \left\langle D^2 \phi_k, D^2 \phi_l \right\rangle_{L_2(T)} = \int_T D^2 \phi_k(t) D^2 \phi_l(t) dt. \tag{10}$$

In the above expressions, $D^n\phi_k(t)$ represents the nth-order differential operator applied to ϕ_k. The solution can be obtained as:

$$\boldsymbol{c} = \left(\boldsymbol{\Phi}^{\boldsymbol{t}}\boldsymbol{\Phi} + \lambda \boldsymbol{R}\right)^{-1}\boldsymbol{\Phi}^{t}\boldsymbol{z}. \tag{11}$$

In the FDA approach, the data from the measurements not only consist of discrete points, but also belong to a smooth function. Typically, this function exhibits some level of smoothness and continuity, where consecutive monitored points z_j and z_{j+1} are close to each other. The analysis is often performed on a group of functional data sets sharing a common performance pattern. In this study, the periodicity was considered by incorporating monthly monitoring.

Functional Bagplot

The concept of functional bagplot extends from the bivariate bagplot developed in [54] that made use of the first two robust scores of the principal component method.

The bivariate bagplot is a variation of the univariate boxplot proposed by Tukey [55]. It represents the semi-space position depth of an element $\theta \in \mathbb{R}^2$ with respect to a bivariate collection of points $Z = \{z_1, z_2, \ldots, z_n\}$. The localization depth of the semi-space $ldepth(\theta, Z), \theta \in \mathbb{R}^2$, $Z = \{z_1, z_2, \ldots, z_n\}$ is defined as the minimum number of z_i included in any closed semi-plane with a boundary line passing through θ. The depth zone D_k is the collection of all points $\theta \in \mathbb{R}^2$, such as $ldepth(\theta, Z) > k; k > 0$. It is necessary to consider that the depth zones are all convex polygons satisfying $D_{k+1} \subset D_k$. The median depth of Z is defined as the θ with the largest $ldepth(\theta, Z)$.

Both the boxplot and the bivariate bagplot have the following components: a central element indicating the median depth, an inner case area, and an outer boundary area. Outliers are individual elements outside these areas. In a bagplot, the element with the highest depth (median depth) is located at the center, and it is surrounded by a case formed by 50% of the elements with the highest depth. This concept is analogous to the interquartile range in a traditional boxplot. By extending the case by a factor, the boundary is determined and any element outside this edge is considered an outlier. The bagplot not only detects outliers, but also provides information about correlation, position, as well as dispersion, skewness, tails, and a confidence zone for the depth median.

Functional bagplots extend the bagplot concept to a collection of curves $\{y_i\}$, $i = 1, 2, \ldots, n$ by utilizing the first two scores obtained through the robust principal components method of y_i [56]:

$$y_i = \mu(x) + \sum_{k=1}^{2} z_{i,k}\varphi_k(x), \tag{12}$$

where $\mu(x)$ represents the mean curve, $\{\phi_k(x)\}$ are the principal components, and $z_i = (z_{i,1}, z_{i,2})$ are the first and second principal component scores. Analyzing the first two principal components enhances the detection of outliers compared with the original functional space [57]. These components typically capture the most important modes of variation [58]. The functional bagplot is particularly useful when the number of variables exceeds the amount of data monitored.

The functional bagplot presents the bivariate bagplot relative to the first and second scores. It includes the median curve, regions with 95% confidence for the median, the 50% case area, and 95% boundary area classified based on depth. Curves outside the 95% convex area are considered functional outliers. Outliers are positioned relative to the boundary in the bagplot.

A detailed development of the bagplot can be found in the study conducted by Ruts and Rousseeuw [59]. The key steps of this procedure are as follows:

1. Calculate the number of elements included in the depth area D_k, denoted as $\#D_k$. Find the index k for which $\#D_k \leq \left(\frac{n}{2}\right) \leq \#D_{k+1}$.

2. Perform a linear interpolation, denoted as $f(x)$, based on the center of mass of the internal area labeled T^* to obtain the case, which is a convex polygon in all cases.
3. Determine the edge by increasing it with respect to T^* by an index of 3. The choice of this quantity is supported by previous findings [54].
4. Mark elements beyond the edge as outliers.

Functional High-Density Region Boxplot

In some studies [56], high-density regions (HDR) are considered a powerful tool to summarize prediction distributions. HDR exhibits versatility by combining modularity and asymmetry in the predictive density. It has the property of being the smallest feasible zone of probability coverage in a sample space at all times. In this article, we focus on a two-dimensional scenario which implies dealing with the minimum surface area.

The boxplot methodology [55] is very useful for studying univariate samples. Hyndman proposed a boxplot methodology based on HDR, which outlines probability distributions and can be extended to n-dimensional multivariate densities [60]. The HDR functional boxplot [56] is a modification of the functional boxplot that arranges the scores $z_i = (z_{i,1}, z_{i,2})$ using a bivariate kernel density approximation.

Let f be a density function and $\{z_i; i = 1, \ldots n\}$ be a random sample of two variables from f. The kernel density approximation for Parzen–Rosenblatt in a bivariate case is as follows [61]:

$$\hat{f}(\boldsymbol{\omega}; a, b) = \frac{1}{nab} \sum_{k=1}^{n} k\left(\frac{\omega_1 - z_{i,1}}{a}\right) k\left(\frac{\omega_2 - z_{i,2}}{b}\right), \tag{13}$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2)'; k(\cdot)$ is a one-dimensional kernel function with symmetry, such that it is verified that $\int k(u)du = 1$, and (a, b), where $a > 0, b > 0$, control the bivariate bandwidth. The HDR boxplot can be defined as:

$$R_\alpha = \left\{ \boldsymbol{z} : \hat{f}(z; a, b) \geq f_\alpha \right\}, \tag{14}$$

where f_α ensures that $Pr(Z \in R_\alpha) \geq 1 - \alpha$. The density estimated within this zone is greater than that outside the zone, with a coverage probability of $(1 - \alpha)$. Notably, the HDR boxplot captures the smallest area in the sample space among all possible areas with a probability coverage of $(1 - \alpha)$, and it includes the mode.

The functional HDR boxplot displays the functional outliers, as well as the mode, determined as $sup_z \hat{f}(\boldsymbol{z}; a, b)$, and the inner and outer areas. The inner area corresponds to the region within the curves of the inner elements of the 50% bivariate HDR boxplot, akin to the inner quartile range of a boxplot. It provides information about the central 50% dispersion of the curves. The outer zone lies outside the curves of the inner elements of the 95% bivariate HDR boxplot and the outliers are the elements not contained within the 95% HDR boxplot.

To address complications arising from numerical integration of $f(x)$, Hyndman proposed a Monte Carlo technique [60]. This involves considering a multivariate random variable X with density $f(x)$ and transforming it into $Y = f(X)$. A new function f_α is obtained by transforming X using the density function, such that $Pr(f(X) \geq f_\alpha) = 1 - \alpha$ holds. Thus, f_α represents the quantile α of Y, and it can be estimated through a sampling process. More details can be found in Reference [62].

3. Results and Discussion

Air quality in the vicinity of the port of El Musel and the western area of Gijón can have a significant impact on people's health and pose a risk factor for various diseases. In this study, we focus on analyzing pollutants based on data collected at the Argentina station, specially targeting pollutants with legislated values for health protection. These pollutants include:

- SO_2—Coal and oil often contain sulfur compounds and their combustion produces sulfur dioxide. Further oxidation of SO_2, usually in the presence of a catalyst such as NO_2, forms H_2SO_4 (sulfuric acid), resulting in acid rain. This is one of the reasons for concern about the environmental impact of the use of these fuels for power generation [32,63]. The limit values set for the protection of human health by European Community Directive 2008/50 [64] are 350 $\mu g/m^3$ over an average period of 1 h and 125 $\mu g/m^3$ over a 24 h mean period.
- NO_2—This gas is considered one of the most important atmospheric pollutants. It is present in the formation of acid rain as HNO_3 (nitric acid) due to reactions with water vapor droplets in the atmosphere [32]. The limit values set in the European legislation [64] are 240 $\mu g/m^3$ over an average period of 1 h and 40 $\mu g/m^3$ for the calendar year.
- CO—Carbon monoxide is a product of the incomplete combustion of fuels such as coal and natural gas. Automobile exhaust is also a source of carbon monoxide [32]. The maximum allowable concentration in air quality legislation [64] is 10 mg per day, based on a daily 8 h mean.
- PM_{10}—Particulate matter in the air represents a complex mixture of organic and inorganic substances. Particulate matter is found in urban areas from thermal power plants, industrial processes, vehicle traffic, domestic coal combustion, and industrial processes and activities [20]. The limit values for PM_{10} collected in [64] are 40 $\mu g/m^3$ per calendar year and 50 $\mu g/m^3$ per day. This daily limit value may not be exceeded more than 35 times per calendar year.
- O_3—Tropospheric ozone of artificial origin is a secondary pollutant, as it is generated from other pollutants in the air due to a series of photochemical reactions of volatile organic compounds and nitrogen oxides [32]. The target value set by [64] legislation for the protection of human health is 120 $\mu g/m^3$ as the daily maximum of the eight-hourly moving averages. This value shall not be exceeded for more than 25 days each calendar year considering a 3 year average.

Our main objective was focused on the analysis of daily PM_{10} measurements, but daily measurements of SO_2, NO, NO_2, CO, and O_3 pollutant gases from the same station were also studied. Since our objective was to extract and analyze statistical information from the collected data, it could happen that the identification of an outlier did not correspond to the consideration of exceeding the limits established by the legislation, since the latter values are fixed and the extreme values come from mathematical analysis and are therefore inherent in the data themselves.

For the pollutants studied, our sample $\left\{x_{ij}^{(k)}\right\}_{j=1}^{n}$ corresponded to $n = 96$ months, spanning from January 2014 to December 2021. In this sample, x_{ij} represents the concentration of each pollutant k measured on day i of month j, with $i = 1, 2, \ldots, 30$ and $j = 1, 2, \ldots, n$, where $k = \{SO_2, NO, NO_2, CO, PM_{10}, O_3\}$.

The results obtained for each of the methods used in the analysis of air pollution at the Argentina station, which was selected to characterize the emissions of the Gijón port area due to its proximity to the El Musel port area in Gijón, are presented below.

Table 1 presents a summary of the outliers identified by the vector analysis and the two methods based on the FDA approach used in this study. It is important to note that the lower outliers were excluded from the analysis results, as they were not considered representative in the context of air pollution and air quality assessment.

Table 2 presents the months and their respective monthly mean values, indicating the months in each year that recorded the highest monthly mean value.

Table 1. Summary of the outliers detected using vector and functional procedures.

Method/Pollutant	SO_2	NO	NO_2	CO	PM_{10}	O_3
Vector			February 2015		March 2014 October 2014 February 2020	September 2018
HDR Boxplot	January 2015 November 2016 December 2016 November 2017 December 2018	November 2014 January 2015 December 2015 December 2016 November 2018	October 2014 February 2015 November 2018 January 2019 March 2020	October 2014 May 2017 November 2017 December 2018 February 2019	Mach 2014 October 2014 March 2015 February 2020 December 2021	May 2018 September 2018 May 2019
Bagplot	March 2014 January 2015 April 2015 February 2016 November 2016 December 2016 February 2017 April 2017 November 2017 November 2018	October 2014 November 2014 February 2015 October 2015 November 2015 December 2015 January 2016 December 2017 February 2018 November 2018 December 2018 February 2019 January 2020 December 2020	October 2014 February 2015 November 2018 January 2019 Mach 2020	October 2014 February 2015 December 2015 December 2016 April 2017 May 2017 November 2017 December 2018 February 2019 October 2021	Mach 2014 October 2014 February 2015 March 2015 February 2020 Mach 2021 December 2021	May 2018 July 2018 September 2018 May 2019 June 2020

Table 2. Month with maximum values as average concentration *.

Pollutant/Year	2014	2015	2016	2017	2018	2019	2020	2021
SO_2 ($\mu g/m^3$)	February (10)	January (17)	December (12)	February (12)	December (13)	February (12)	January (8)	December (8)
NO ($\mu g/m^3$)	November (18)	December (30)	December (30)	January (20)	December (25)	December (16)	January (19)	December (17)
NO_2 ($\mu g/m^3$)	November (34)	January (38)	December (40)	January (37)	November (37)	February (38)	January (31)	December/January (27)
CO (mg/m^3)	October (0.64)	December (0.68)	December (0.75)	May (0.86)	December (0.83)	February (0.93)	October (0.65)	January (0.50)
PM10 ($\mu g/m^3$)	October (47)	December (44)	December (36)	October (35)	December (40)	February (42)	February (46)	December (49)
O_3 ($\mu g/m^3$)	July/May (46)	July (46)	April (47)	April (54)	September (65)	April (58)	April (53)	June/April (53)

* The average concentration is given in brackets. Source: adapted from [65].

3.1. Vector Analysis Results

During the data exploration phase, it was observed that the data set did not follow a normal distribution. An Anderson–Darling test was performed, yielding a p-value of less than 0.005 for all pollutants. Therefore, the normality of the data was not accepted at the 5% significance level.

A Box–Cox transformation was applied to the data to obtain a normal distribution. This transformation was appropriate because the pollutant values were always positive and there were no observations of pollutants equal to zero during the study period.

Table 3 shows the optimal λ value used for the transformation along the upper and lower limits of the 95% confidence interval for each pollutant.

Table 3. Optimal values of λ for Box–Cox Transformation for SO_2, NO, NO_2, CO, PM_{10}, and O_3 pollutants.

Pollutant	Lower Confidence Limit	Upper Confidence Limit	Optimal λ
SO_2	−0.42	−0.29	−0.35
NO	−0.11	−0.02	−0.07
NO_2	0.26	0.40	0.33
CO	−0.44	−0.31	−0.38
PM_{10}	0.06	0.23	0.15
O_3	0.78	0.94	0.86

The generalized ESD (extreme deviation studied) test was conducted as described above. A value of $r = 1$ for all pollutants, and for the pollutant PM_{10}, the test was repeated with $r = 2$ after having detected one outlier with $r = 1$. The results obtained are presented in Table 4. The Rosner test identified only one outlier for PM_{10} in this study.

Table 4. Results of generalized ESD test.

Pollutant	i	Mean (i)	Standard Deviation (i)	Value	Observation (Day)	R (i + 1)	λ (i + 1)	Outlier
SO_2	0	0.57981	0.11723	1.00000	2673	3.58428	4.28987	FALSE
NO	0	0.89639	0.04345	0.76782	665	2.95901	4.28987	FALSE
NO_2	0	2.80590	0.45107	4.28506	1365	3.27923	4.28987	FALSE
CO	0	1.58098	0.35121	0.69273	1796	2.52911	4.28987	FALSE
PM_10	0	1.62088	0.09149	2.02692	2611	4.43804	4.28987	TRUE
	1	1.62074	0.09119	1.26132	446	3.94134	4.28979	FALSE
O_3	0	21.92891	7.14035	46.24131	1709	3.40493	4.28987	FALSE

3.2. Results of the FDA Approach

By applying the described smoothing technique to the initial sample $\left\{x_{ij}^{(k)}\right\}_{j=1}^{n}$, a new sample $\{x_j(t)\}$ was generated for each analyzed pollutant k. In this new sample x_j represents a new basis function derived from Fourier basis functions. In this analysis, a set of 100-element Fourier basis functions was specifically utilized. The correlation coefficient between the discrete values of the initial sample and the values of the functions generated from the new sample was 0.99, indicating a strong 99% correlation between the functional sample and the discrete values.

The FDA approach was performed after the samples were constructed using the functional smoothing method. Specifically, the two functional methods mentioned in the previous section were applied to a data set obtained from the Argentina air quality monitoring station.

Figures 2 and 3 show the functional bagplot and the HDR boxplot, respectively, for the data collected at the Argentina station.

Both the functional bagplot and the HDR boxplot methods provide statistical means to identify outliers. Outliers are identified as curves that lie outside the light gray shaded areas that represent the bands corresponding to 95% confidence.

Table 1 shows the outliers obtained using the functional bagplot and HDR boxplot methods. Each outlier indicates the month in which a particular pollutant had a concentration outside the 95% confidence bands during the 8-year study period. The lower outliers were excluded from Table 1 as they were not considered representative for the purpose of the pollutant concentration analysis.

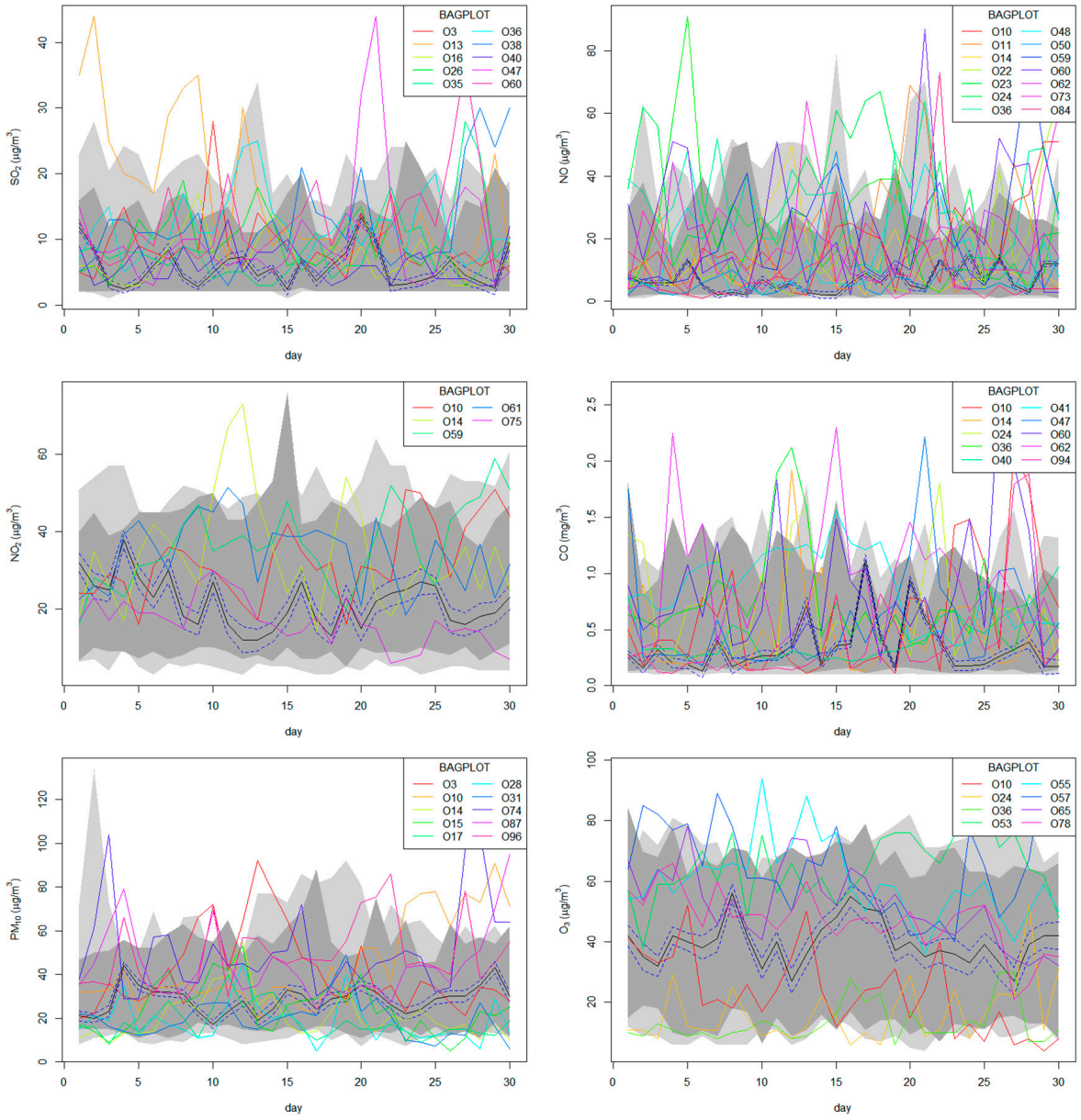

Figure 2. Functional bagplot analysis of the pollutant emissions of SO_2, NO, NO_2, CO, PM_{10}, and O_3 at Argentina air quality monitoring station in Gijón. O_j indicates an outlier in month j, where $j = 1$ represents January 2014 and $j = 96$ represents December 2021.

The highest values were recorded in the autumn and winter months for SO_2, NO, NO_2, CO, and PM_{10} air pollutants. This is in agreement with observations made in several studies [66–68]. However, some exceptions can be seen, as in 2017 when the month with the highest monthly average concentration for the pollutant CO was May, with a value of 0.86 mg/m^3, which corresponded to the highest value in the series after February 2019, with a value of 0.93 mg/m^3.

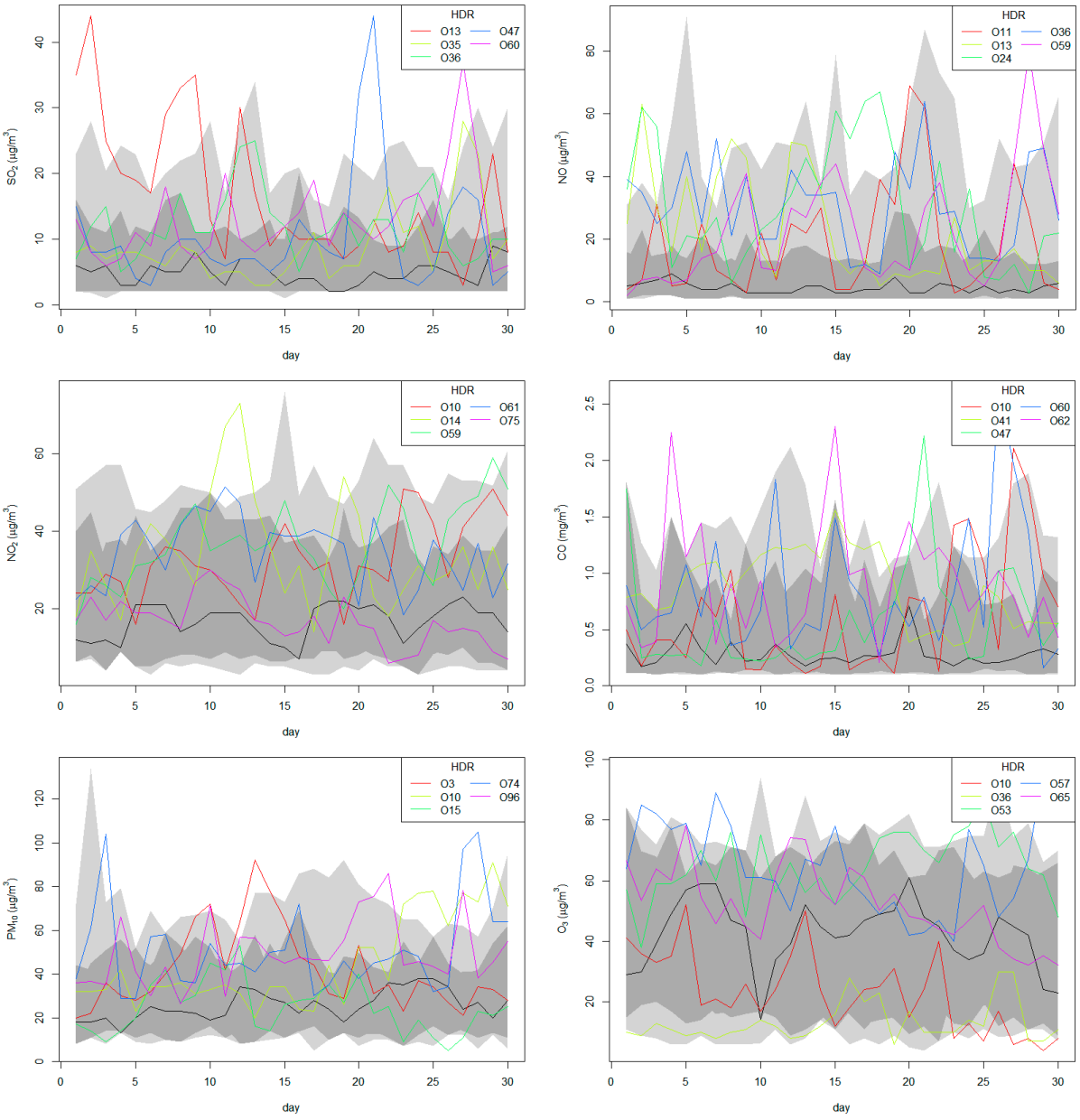

Figure 3. HDR boxplot analysis of the pollutant emissions of SO_2, NO, NO_2, CO, PM_{10}, and O_3 at Argentina air quality monitoring station in Gijón. O_j indicates an outlier in month j, where $j = 1$ represents January 2014 and $j = 96$ represents December 2021.

By observing the graph of the daily concentration averages for the pollutant CO corresponding to the month of May for each of the 8 years studied (2014 to 2021) in Figure 4, it can be observed that the month of May presented an atypical behavior with values much higher than the rest of the annual series for the same month.

On the other hand, the pollutant O_3 presented the highest monthly mean values in spring and summer, as observed in some studies as [69], which correspond to the periods of highest solar activity. However, an exception can be considered in 2018, in which the month with the highest average concentration was September.

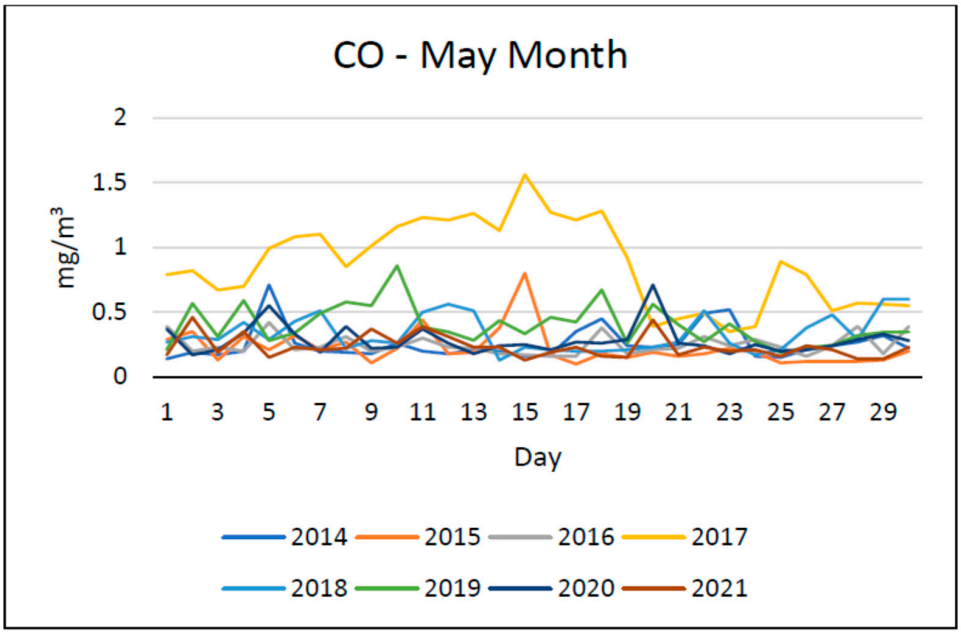

Figure 4. Average daily concentration of the pollutant CO during the month of May in the 8-year study period (2014–2021).

Regarding climatology, September 2018, as shown in Figure 5, was the third month with the most hours of sunshine with 183 h, compared with 251 in August and 214 in May. However, September 2018 was the month with the lowest precipitation month in 2018, with only 50.3 mm, following an atypical December with only 45.1 mm. The high sunshine hour activity combined with low precipitation in September 2018 could explain why the annual maximum of monthly averages was delayed until September in that year.

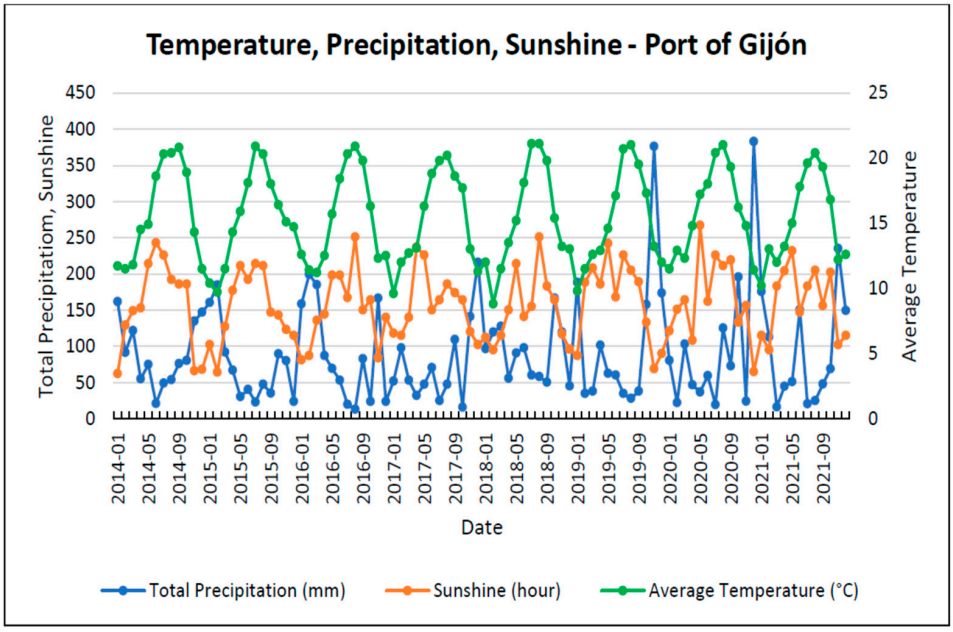

Figure 5. Temperature, precipitation, and sunshine hours chart in Gijón Port during the 8-year study period (2014–2021).

With vector analysis using the Z-score method, no outliers were detected for SO_2, NO, and CO pollutants. Only one outlier was detected for NO_2 corresponding to February 2015; three for PM_{10} in March 2014, October 2014, and February 2020; and one outlier for O_3 in September 2018. Note that lower outliers were not considered as they were not

representative in this air pollution study. The number of lower outliers detected using this method was PM_{10} (five) and NO_2 (two). The Rosner test only detected one outlier for PM_{10} in March 2021.

The number of outliers detected using the HDR boxplot and bagplot methods of the FDA approach was considerable, especially in the case of the bagplot. Outliers were detected for all contaminants using the FDA approach methods. Table 1 shows the outliers obtained by the vector and methods based on the FDA approach.

Based on the results obtained, it was observed that the outliers detected using vector analysis were also identified with the methods based on the FDA approach. This observation suggests that non-normality of the recorded measurements' data set may be the reason for their presence. Classical methods are highly robust when dealing with data from normal distributions; however, their efficiency decreases when confronted with non-normal distributions.

Vector methods consider point observations to study the data. However, because air pollution from gases does not depend on single measurements, but on a series of measurements that may exceed the limits set by legislation within a specific time interval or, in this case, the statistically established limits derived from the data set, the FDA approach treats these observations as functional data rather than individual points. As a result, our approach involved treating air pollutant gas measurements as a collection of curves rather than as vectors, which can better capture variability. This perspective allowed us to view the original data set as a time-dependent function along a continuous domain rather than a set of discrete values at different points in time.

Methods based on the FDA approach are also more sensitive to detecting outliers than vector methods, especially when the data have complex or non-linear patterns. The FDA approach allows individual curves to be fitted to the data, which can be useful for detecting specific outliers in each curve.

However, they have some limitations, such as the fact that methods based on the FDA approach may require a higher level of knowledge and expertise compared with classical vector methods, coupled with higher computational complexity, which may increase the time and resources needed to perform the analysis. In the situation of data sets that follow a normal distribution, classical methods are generally quite robust and simpler. However, such classical methods lose efficiency on data sets that do not follow a normal distribution.

The proposed methods can be sensitive to the choice of the functional representation of the data and to the dimension reduction techniques. The choice of appropriate basis functions or dimension reduction methods can affect the quality of the functional representation and the subsequent detection of outliers.

Bias may also be introduced by the presence of incomplete or missing data. The methods assume complete and reliable observations, but in the real world, data may be missing for a variety of reasons, such as sensor failure or maintenance issues. Missing data can introduce bias and affect the accuracy of outlier detection.

As mentioned above, one advantage of methods based on the FDA approach over the identification of outliers by classical, i.e., vector methods, is the absence of need to assume normality. Therefore, there is no requirement to test for normality using methods such as Anderson–Darling or to perform a Box–Cox transformation.

In this study, the HDR boxplot method performed worse in detecting outliers of both methods based on the FDA approach. All outliers identified by the HDR boxplot method were detected by the bagplot method, except for two values for the pollutant NO, corresponding to months 13 (January 2015) and 36 (December 2016).

Figure 6 displays the graphs depicting the functions obtained after smoothing the pollutant emission data for SO_2, NO, NO_2, CO, PM_{10}, and O_3. Each pollutant had 96 functions representing each month of the 8-year study period conducted at the air quality monitoring station in Gijón.

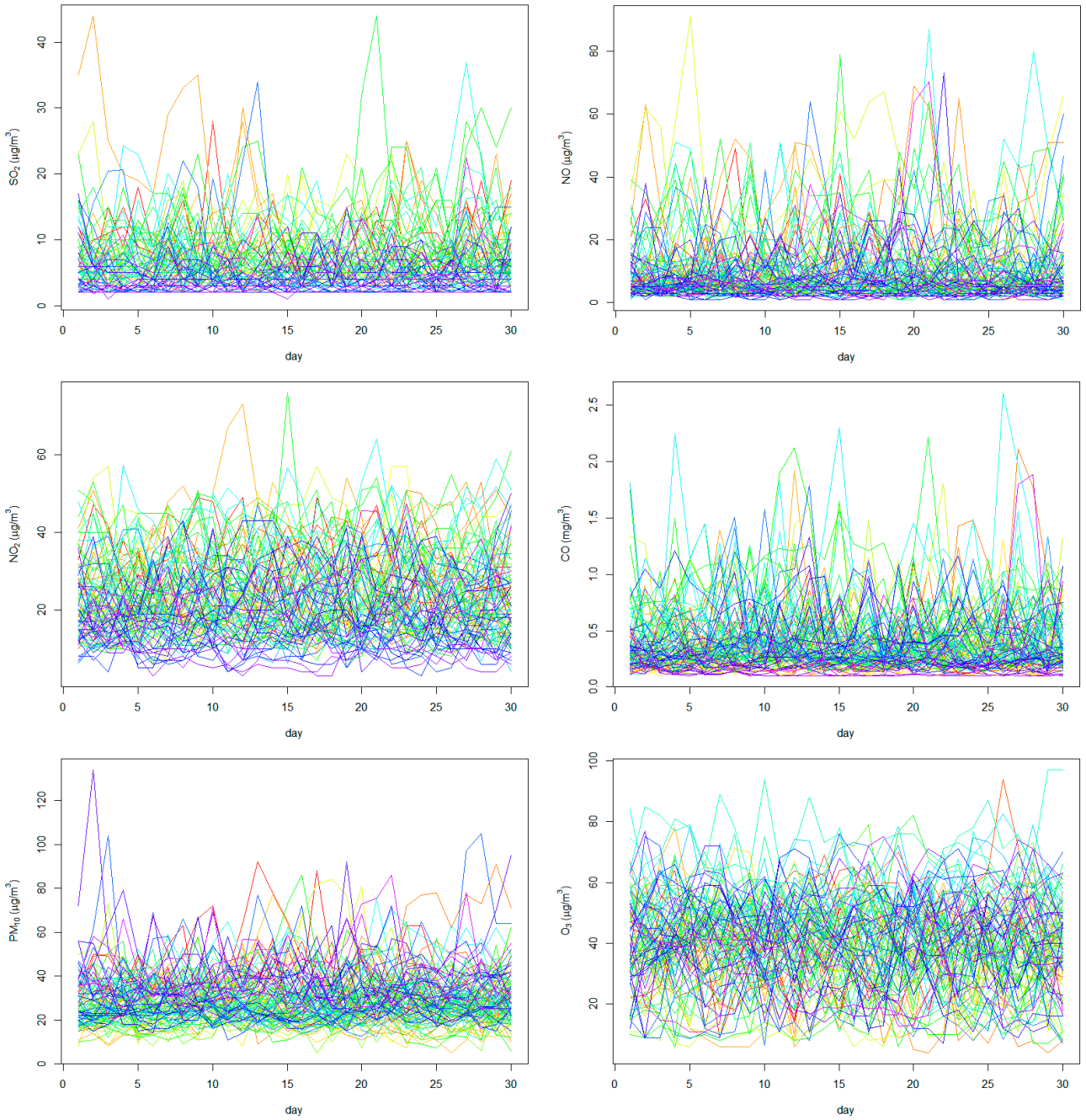

Figure 6. Pollutant emission functions of SO_2, NO, NO_2, CO, PM_{10}, and O_3 at Argentina air quality monitoring station in Gijón during the 8-year study period.

With regard to maritime traffic, as indicated in the internal report of the Gijon Port Authority [70], the Port of El Musel had a significant decrease in the movement of goods through its facilities in 2020. It was a year marked by the health crisis and, in the case of the pier of Gijón, by the strong impact of decarbonization, as well as by the productive adjustment of steel production in the Asturian region. A slight recovery can be seen in 2021. Table 5 shows the evolution of maritime traffic in terms of number of vessels, gross tonnage, and dry bulk from 2010 to 2021.

Table 5. Port maritime traffic in Gijón Port (2010–2021).

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Number of ships	1050	1065	1125	1251	1315	1183	1156	1298	1229	1187	1127	1189
Gross Tonnage (thousands of tons)	14,254	17,735	19,091	20,504	21,192	19,097	18,311	19,607	18,226	18,005	16,468	16,569 9246
Dry Bulk (thousands of tons)	10,602	10,256	12,024	1669	12,382	15,355	12,036	14,768	12,717	9790	7010	2021 1189
Number of ships Gross Tonnage	1050	1065	1125	1251	1315	1183	1156	1298	1229	1187	1127	16,569 9246

Source: adapted from [31,71].

Moldanová et al. [72] presented a PM emission factor from diesel-powered ships of 5.3 g, with a bimodal particle size distribution, one of the maxima being in the diameter of 7 microns and the other at 0.5 microns. Another study found that the total PM emission was 36 tons/year at Ambarli Port [73] and other researchers estimated 57.4 tons of PM per year for 7520 ships in the Candali Gulf [74], both located in Turkey.

According to Mueller et al. [75], factors such as road traffic, industrial and residential emissions, as well as ships' emissions can affect air quality in port cities. The Argentina station in Gijón, due to its location, can receive not only emissions from the port of El Musel, but also emissions from the industrial areas to the west of Gijón and emissions from traffic in the area, as well as residential emissions. Another important factor is the daily truck traffic generated by the Port of El Musel. According to the consulting firm Vectio Traffic & Transport Planning (https://www.vectio.es) in a report prepared for the City Council of Gijón, the average number of trucks passing through one of the main access roads to the port was 1052 per day.

For the entire period 2004 to 2021, the Argentina station recorded exceedances of the limit values only in the case of the pollutant PM_{10}. Table 6 shows the number of exceedance days in that period and the annual average concentration of PM_{10}. The concentration of the others air pollutants were below the legal limits. Between the years 2009 and 2013, more than 35 exceedances per year of the limit value for the daily average concentration of PM_{10}, which is set at 50 $\mu g/m^3$, were reached. In addition, there is the possibility of receiving pollutant emissions from several sources, as indicated above. These reasons have led to the implementation of an action plan to improve air quality.

Table 6. Annual average of PM10 at the air quality station Argentina and number of exceedance days during the period 2004–2021.

Year	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21
Exceedance days	76	93	84	70	25	45	41	63	44	41	36 *	21	6	14	16	11	26	38 *
Annual average ($\mu g/m^3$)	48	43	42	39	31	36	33	37	34	35	33	28	25	29	31	28	31	38

(*) Due to discounts for Saharan intrusions, there were no instances of legislative non-compliance regarding exceedance days in 2014 and 2021. Exceeded values are highlighted in red. It is important to note that the current legislative values were revised in 2004 and have remained unchanged since then.

For these reasons, the Government of the Principality of Asturias and the Gijón City Council commissioned a study from the Carlos III Institute of Health, which serves as the National Reference Laboratory for Air Quality. This study [76] recommended a new station located further west to better characterize industrial emissions.

Since the implementation of the actions in the 2014 Air Quality Improvement Plan by the Government of the Principality of Asturias and the City Council of Gijón, there has been a continuous decrease in the number of daily exceedances of PM_{10} at the Argentina station (see Table 6), a trend that was broken in 2017 and 2018 with 14 and 16 daily exceedances, and subsequently in 2020 and 2021 with 26 and 38 exceedances, respectively. These data

should be taken with caution, since at the end of this study period, the economic activities of shipping and industry were paralyzed or slowed down due to COVID-19, but there was an increase in the annual average PM_{10} values, as well as in the number of exceedance days.

In 2021, the last year with complete data for this study, the Argentina station was in compliance with the number of allowable days exceeding the daily limit for PM_{10}, after applying discounts for Saharan intrusions. Saharan intrusions are a source of dust from the Sahara Desert [77]. Table 7 shows the annual mean and standard deviation for each of the pollutants for the 8-year study period (2014–2021) at the Argentina air quality station.

Table 7. Average annual value and standard deviation * of each pollutant at the Argentina air quality station during the period 2004–2021.

Year	SO_2 ($\mu g/m^3$)	NO ($\mu g/m^3$)	NO_2 ($\mu g/m^3$)	CO (mg/m^3)	PM_{10} ($\mu g/m^3$)	O_3 ($\mu g/m^3$)
2014	7.06 (1.99)	9.92 (5.31)	25.68 (6.95)	0.35 (0.12)	33.01 (7.49)	38.60 (8.72)
2015	7.23 (3.71)	12.21 (8.63)	27.29 (8.00)	0.35 (0.16)	28.38 (6.81)	35.75 (10.15)
2016	7.13 (2.10)	9.50 (7.79)	21.68 (8.48)	0.39 (0.16)	25.30 (4.64)	36.28 (9.55)
2017	8.11 (2.75)	9.27 (5.43)	26.25 (7.39)	0.50 (0.15)	28.85 (3.77)	38.92 (8.70)
2018	6.51 (3.30)	9.45 (7.20)	25.13 (7.53)	0.43 (0.17)	30.45 (4.81)	48.66 (12.31)
2019	5.49 (2.26)	10.54 (3.64)	22.75 (8.17)	0.48 (0.15)	27.73 (5.48)	44.25 (8.93)
2020	4.01 (1.76)	7.39 (5.20)	18.21 (7.15)	0.38 (0.15)	31.13 (9.26)	39.72 (10.18)
2021	4.37 (1.62)	8.47 (4.16)	18.53 (6.76)	0.27 (0.12)	38.06 (6.41)	40.56 (9.60)

(*) The value without brackets indicates the annual mean value and the value in brackets indicates the standard deviation.

As for the other air pollutants, all of them showed a decrease in 2020 compared with the values obtained in 2019. In 2021, the last year of the study, the annual average concentration of SO_2 increased slightly from 4 to 5 $\mu g/m^3$. The same happened for NO, which increased from 7 to 9 $\mu g/m^3$; with NO_2, from 18 to 19 $\mu g/m^3$; and O_3 remained constant at 40 $\mu g/m^3$ from 2020 to 2021. In contrast, CO continued the downward trend from 0.38 mg/m^3 in 2020 to 0.27 mg/m^3 in 2021. In 2021, there were no exceedances of the hourly limit value for any of the studied air pollutants.

The FDA approach for outlier detection has been used in other works, although with different approaches, either in the methods based on the FDA approach or in the field of application, such as those Díaz Muñiz et al. [78] applied to water quality control, in the emissions of polluting gases affecting urban areas [53]. Sguera et al. studied NOx levels in an industrial area to determine the variation in pollution between working days and weekends or holidays [28]. Chiou et al. [79] used an FDA approach to detect outliers in traffic flow data. In another study, the HDR boxplot and bagplot methods were used for the detection of outliers in data with hydrological applications [80]. In a recent study, new methods based on an FDA approach were proposed for identifying outliers in functional data sets. They were applied to climatological data series in Spain [81].

The calculations were performed on an AMD Ryzen 7 5800H processor, 3.20 GHz and 32 GB of RAM. The computational cost, using PM_{10} as an example, was 4.01 s for the HDR and 0.08 s for the bagplot, compared with 0.02 s for the Z-score and less than 0.01 for the Rosner test.

4. Conclusions

In this study, we analyzed a set of data from measurements recorded by the air quality station of the Government of the Principality of Asturias, located closest to the Port of El Musel (Gijón), in order to detect outliers. This analysis was carried out because of the social alarm caused by the pollution episodes, especially PM_{10}, detected in this area, as an aid to decision making by local authorities, and for a possible means of early warning.

We used a vector method included in the European standards, such as the Z-score, a test to detect outliers, such as the generalized ESD test and the Rosner test, and more recent techniques to treat the discrete values of the observations as a series of continuous functions as a function of time.

Outlier detection using methods based on the FDA approach, such as HDR boxplot and bagplot, was more powerful than vector methods, such as Z-score and the generalized ESD test.

On the one hand, the analysis of the observations collected between 2014 and 2021 has shown the presence of the following upper outliers of air concentration for the vector method: NO_2 (one outlier), PM_{10} (three), and O_3 (one). No outliers were detected for the pollutants SO_2, NO, and CO. The generalized ESD test detected a single outlier for PM_{10} and none for the other pollutants.

The functional bagplot method detected a greater number of outliers than the HDR boxplot, with the following number of outliers being higher for the HDR boxplot method: SO_2 (five), NO (five), NO_2 (five), CO (five), PM_{10} (five), and O_3 (three). The bagplot method identified SO_2 (10), NO (13), NO_2 (5), CO (10), PM_{10} (7), and O_3 (5).

According to the results obtained in the pollutant analysis, the methods based on the FDA approach allowed us to retrieve more information than the vector methods. The latter compare averages, but do not take into account temporal variations.

Methods based on the FDA approach, in addition to having detected a greater number of outliers, have the advantage of not requiring that the data come from a normal distribution and, therefore, no hypothesis was made about the distribution followed by the recorded observations. This avoided the need to perform transformations on the data. In addition, this methodology considers the temporal correlation structure.

As for the origin of the pollution at the Argentina station, it could have several sources. In addition to the port of El Musel, it may have an industrial and traffic origin. For this reason, a study commissioned by the City Council of Gijón and the Government of the Principality of Asturias determined the need for a new station in a more western location. Therefore, new studies incorporating observations from this new station are needed to determine the origin.

Future research work, in addition to considering this new station, should focus on the application of this FDA approach methodology to statistical study and introduce sweep rules to evaluate trends in the functional samples.

In general, some of the factors that cause pollution to increase in an area are economic and population growth. In times of economic recession, pollution tends to decrease. It is possible to apply this air pollution outlier detection methodology to other air quality stations and in other geographic areas with success. It is also possible to apply this technique to a wide range of other types of problems, not necessarily related to air pollution, but in other areas such as water quality in water bodies, etc., to detect outliers. However, the specific characteristics of the site must always be taken in account.

This functional outlier detection methodology can be used as an early warning tool that allows authorities to take preventive decisions and effective measures to preserve the health of citizens.

Author Contributions: Conceptualization, L.A.M.-G., P.J.G.-N., F.S.L. and L.Á.-d.-P.; methodology, P.J.G.-N., E.G.-G. and A.B.-S.; software, L.A.M.-G., E.G.-G., F.S.L., A.B.-S. and L.Á.-d.-P.; validation, L.A.M.-G., E.G.-G., A.B.-S. and L.Á.-d.-P.; formal analysis, L.A.M.-G., E.G.-G., P.J.G.-N., F.S.L., A.B.-S. and L.Á.-d.-P.; investigation, L.A.M.-G., E.G.-G., P.J.G.-N., F.S.L., A.B.-S. and L.Á.-d.-P.; resources, L.A.M.-G., P.J.G.-N., F.S.L. and A.B.-S.; data curation, L.A.M.-G., P.J.G.-N., F.S.L., A.B.-S. and L.Á.-d.-P.; writing—original draft preparation, L.A.M.-G., P.J.G.-N. and L.Á.-d.-P.; writing—review and editing, E.G.-G., F.S.L. and A.B.-S.; visualization, L.A.M.-G. and P.J.G.-N.; supervision, A.B.-S. and L.Á.-d.-P.; project administration, E.G.-G. and F.S.L.; funding acquisition, L.Á.-d.-P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Universidad de León (Spain). Funding number: Program 463A.3.01

Data Availability Statement: The datasets generated and/or analyzed in the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, L.K.; Pereira, N.C.; Hung, Y.-T. *Air Pollution Control Engineering*; Humana Press: Totowa, NJ, USA, 2004; ISBN 978-1-58829-161-5.
2. Profillidis, V.A.; Botzoris, G.N. Evolution and Trends of Transport Demand. In *Modeling of Transport Demand*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 47–87. ISBN 978-0-12-811513-8.
3. United Nations Conference on Trade and Development. *Review of Maritime Transport 2017*; United Nations Publication: Geneva, Switzerland, 2017; ISBN 978-92-1-362808-9.
4. Noya Arquero, F.J. New Port Facilities at Punta Langosteira, Spain. *Proc. Inst. Civ. Eng. Marit. Eng.* **2008**, *161*, 101–106. [CrossRef]
5. Doldán-García, X.R.; Chas-Amil, M.L.; Touza, J. Estimating the Economic Impacts of Maritime Port Development: The Case of A Coruña, Spain. *Ocean Coast. Manag.* **2011**, *54*, 668–677. [CrossRef]
6. Samanés, T.; González-Cancelas, M.N.; Serrano, B.M.; Corral, M.M. Determination of Key Operational and Environmental Performance Indicators in Dirty Solid Bulk Terminals. Implementation in the Spanish Port System. *World Sci. News* **2019**, *132*, 16–34.
7. Sorte, S.; Rodrigues, V.; Borrego, C.; Monteiro, A. Impact of Harbour Activities on Local Air Quality: A Review. *Environ. Pollut.* **2020**, *257*, 113542. [CrossRef]
8. Cullinane, K. *Targeting the Environmental Sustainability of European Shipping the Need for Innovation in Policy and Technology*; Technical Report; EPSD—Europoean Panel of Sustainable Development: Gothenburg, Sweden, 2014. [CrossRef]
9. Isakson, J.; Persson, T.A.; Selin Lindgren, E. Identification and Assessment of Ship Emissions and Their Effects in the Harbour of Göteborg, Sweden. *Atmos. Environ.* **2001**, *35*, 3659–3666. [CrossRef]
10. Lack, D.A.; Corbett, J.J. Black Carbon from Ships: A Review of the Effects of Ship Speed, Fuel Quality and Exhaust Gas Scrubbing. *Atmos. Chem. Phys.* **2012**, *12*, 3985–4000. [CrossRef]
11. Clemente, Á.; Yubero, E.; Galindo, N.; Crespo, J.; Nicolás, J.F.; Santacatalina, M.; Carratala, A. Quantification of the Impact of Port Activities on PM10 Levels at the Port-City Boundary of a Mediterranean City. *J. Environ. Manag.* **2021**, *281*, 111842. [CrossRef]
12. Artíñano, B.; Gómez-Moreno, F.J.; Pujadas, M.; Moreno, N.; Alastuey, A.; Querol, X.; Martín, F.; Guerra, A.; Luaces, J.A.; Basora, J. Measurement of Particulate Concentrations Produced during Bulk Material Handling at the Tarragona Harbor. *Atmos. Environ.* **2007**, *41*, 6344–6355. [CrossRef]
13. Alastuey, A.; Moreno, N.; Querol, X.; Viana, M.; Artíñano, B.; Luaces, J.A.; Basora, J.; Guerra, A. Contribution of Harbour Activities to Levels of Particulate Matter in a Harbour Area: Hada Project-Tarragona Spain. *Atmos. Environ.* **2007**, *41*, 6366–6378. [CrossRef]
14. Almeida, S.M.; Silva, A.V.; Freitas, M.C.; Marques, A.M.; Ramos, C.A.; Silva, A.I.; Pinheiro, T. Characterization of Dust Material Emitted during Harbour Activities by K0-INAA and PIXE. *J. Radioanal. Nucl. Chem.* **2012**, *291*, 77–82. [CrossRef]
15. Toscano, D.; Murena, F. Atmospheric Ship Emissions in Ports: A Review. Correlation with Data of Ship Traffic. *Atmos. Environ.* **2019**, *4*, 100050. [CrossRef]
16. González, Y.; Rodríguez, S.; Guerra García, J.C.; Trujillo, J.L.; García, R. Ultrafine Particles Pollution in Urban Coastal Air Due to Ship Emissions. *Atmos. Environ.* **2011**, *45*, 4907–4914. [CrossRef]
17. Thera, B.T.P.; Dominutti, P.; Öztürk, F.; Salameh, T.; Sauvage, S.; Afif, C.; Çetin, B.; Gaimoz, C.; Keleş, M.; Evan, S.; et al. Composition and Variability of Gaseous Organic Pollution in the Port Megacity of Istanbul: Source Attribution, Emission Ratios, and Inventory Evaluation. *Atmos. Chem. Phys.* **2019**, *19*, 15131–15156. [CrossRef]
18. Anjos, M.; Lopes, A. Sea Breeze Front Identification on the Northeastern Coast of Brazil and Its Implications for Meteorological Conditions in the Sergipe Region. *Theor. Appl. Climatol.* **2019**, *137*, 2151–2165. [CrossRef]
19. Saraga, D.E.; Tolis, E.I.; Maggos, T.; Vasilakos, C.; Bartzis, J.G. PM2.5 Source Apportionment for the Port City of Thessaloniki, Greece. *Sci. Total Environ.* **2019**, *650*, 2337–2354. [CrossRef]
20. Pérez, N.; Pey, J.; Reche, C.; Cortés, J.; Alastuey, A.; Querol, X. Impact of Harbour Emissions on Ambient PM10 and PM2.5 in Barcelona (Spain): Evidences of Secondary Aerosol Formation within the Urban Area. *Sci. Total Environ.* **2016**, *571*, 237–250. [CrossRef]
21. Soggiu, M.E.; Inglessis, M.; Gagliardi, R.V.; Settimo, G.; Marsili, G.; Notardonato, I.; Avino, P. PM10 and PM2.5 Qualitative Source Apportionment Using Selective Wind Direction Sampling in a Port-Industrial Area in Civitavecchia, Italy. *Atmosphere* **2020**, *11*, 22. [CrossRef]
22. Fileni, L.; Mancinelli, E.; Morichetti, M.; Passerini, G.; Rizza, U.; Virgili, S. Air Pollution in Ancora Harbour, Italy. In Proceedings of the 2019 International Conference on Maritinme Transport, Rome, Italy, 10–12 September 2019; pp. 199–208.
23. Wang, X.; Shen, Y.; Lin, Y.; Pan, J.; Zhang, Y.; Louie, P.K.K.; Li, M.; Fu, Q. Atmospheric Pollution from Ships and Its Impact on Local Air Quality at a Port Site in Shanghai. *Atmos. Chem. Phys.* **2019**, *19*, 6315–6330. [CrossRef]
24. Deneshkumar, V.; Senthamaraikannan, K.; Senthamaraikannan, M. Identification of Outliers in Medical Diagnostic System Using Data Mining Techniques. *Int. J. Stat. Appl.* **2014**, *4*, 241–248. [CrossRef]
25. Zou, M.; Djokic, S.Z. A Review of Approaches for the Detection and Treatment of Outliers in Processing Wind Turbine and Wind Farm Measurements. *Energies* **2020**, *13*, 4228. [CrossRef]
26. Araki, S.; Shimadera, H.; Yamamoto, K.; Kondo, A. Effect of Spatial Outliers on the Regression Modelling of Air Pollutant Concentrations: A Case Study in Japan. *Atmos. Environ.* **2017**, *153*, 83–93. [CrossRef]
27. O'Leary, B.; Reiners, J.J.; Xu, X.; Lemke, L.D. Identification and Influence of Spatio-Temporal Outliers in Urban Air Quality Measurements. *Sci. Total Environ.* **2016**, *573*, 55–65. [CrossRef] [PubMed]

28. Sguera, C.; Galeano, P.; Lillo, R.E. Functional Outlier Detection by a Local Depth with Application to NOx Levels. *Stoch. Environ. Res. Risk Assess.* **2016**, *30*, 1115–1130. [CrossRef]
29. Carantoña Alvarez, F. (Ed.) . *El Musel: Historia de Un Puerto*; Lunwerg Editores: Barcelona, Spain, 2005; ISBN 978-84-9785-184-8.
30. Gijon Port Authority Port of Gijon Annual Report 2019. Available online: https://www.puertogijon.es/wp-content/uploads/2020/08/Memoria-2019.pdf (accessed on 17 April 2022).
31. Gijon Port Authority Port of Gijon Annual Report 2020. Available online: https://www.puertogijon.es/wp-content/uploads/2021/09/Memoria-Anual-Puerto-Gijon-2020.pdf (accessed on 26 October 2022).
32. Kiely, G. *Environmental Engineering*; Trwin/McGraw-Hill: Boston, MA, USA, 1998; ISBN 978-0-07-116424-5.
33. Schweitzer, P.E.; Philip, A. *Fundamentals of Metallic Corrosion: Atmospheric and Media Corrosion of Metals*; CRC Press: Boca Raton, FL, USA, 2006; ISBN 978-0-429-12713-7.
34. Preradovic, L.; Ilic, P.; Markovic, S.; Janjus, Z. Meteorological Parameters and Pollution Caused by Sulfur Dioxide and Their Influence on Construction Materials and Heritage. *Facta Univ. Electron. Energ.* **2011**, *24*, 9–20. [CrossRef]
35. Jacobson, M.Z. *Fundamentals of Atmospheric Modeling*, 2nd ed.; Cambridge University Press: New York, NY, USA, 2005; ISBN 978-0-521-83970-9.
36. Thurston, G.D.; Ito, K. Epidemiological Studies of Acute Ozone Exposures and Mortality. *J. Expo. Sci. Environ. Epidemiol.* **2001**, *11*, 286–294. [CrossRef]
37. McDonnell, W.F.; Abbey, D.E.; Nishino, N.; Lebowitz, M.D. Long-Term Ambient Ozone Concentration and the Incidence of Asthma in Nonsmoking Adults: The Ahsmog Study. *Environ. Res.* **1999**, *80*, 110–121. [CrossRef]
38. Hassanvand, M.S.; Naddafi, K.; Kashani, H.; Faridi, S.; Kunzli, N.; Nabizadeh, R.; Momeniha, F.; Gholampour, A.; Arhami, M.; Zare, A.; et al. Short-Term Effects of Particle Size Fractions on Circulating Biomarkers of Inflammation in a Panel of Elderly Subjects and Healthy Young Adults. *Environ. Pollut.* **2017**, *223*, 695–704. [CrossRef]
39. Chauhan, A.J.; Johnston, S.L. Air Pollution and Infection in Respiratory Illness. *Br. Med. Bull.* **2003**, *68*, 95–112. [CrossRef]
40. Ghosh, R.; Joad, J.; Benes, I.; Dostal, M.; Sram, R.J.; Hertz-Picciotto, I. Ambient Nitrogen Oxides Exposure and Early Childhood Respiratory Illnesses. *Environ. Int.* **2012**, *39*, 96–102. [CrossRef]
41. Arroyave, C.; Morcillo, M. The Effect of Nitrogen Oxides in Atmospheric Corrosion of Metals. *Corros. Sci.* **1995**, *37*, 293–305. [CrossRef]
42. Osborne, J. Notes on the Use of Data Transformations. *Pract. Assess. Res. Eval.* **2002**, *8*, 1–7. [CrossRef]
43. Box, G.E.P.; Cox, D.R. An Analysis of Transformations. *J. R. Stat. Soc. Ser. B Methodol.* **1964**, *26*, 211–243. [CrossRef]
44. Peña Sánchez de Rivera, D.; Peña Sánchez de Rivera, J.I. Un Contraste de Normalidad Basado En La Transformación de Box-Cox. *Estadística Española* **1986**, *110*, 33–46.
45. Anderson, T.W.; Darling, D.A. A Test of Goodness of Fit. *J. Am. Stat. Assoc.* **1954**, *49*, 765–769. [CrossRef]
46. Aldás Manzano, J.; Uriel Jiménez, E. *Análisis Multivariante Aplicado con R*, 2nd ed.; Paraninfo Thomson Learning: Madrid, Spain, 2017; ISBN 978-84-283-2969-9.
47. *ISO/IEC 17043:2010*; Conformity Assessment—General Requirements for Proficiency Testing. ISO: Geneva, Switzerland, 2010; p. 51.
48. Grubbs, F.E. Sample Criteria for Testing Outlying Observations. *Ann. Math. Statist.* **1950**, *21*, 27–58. [CrossRef]
49. Rosner, B. Percentage Points for a Generalized ESD Many-Outlier Procedure. *Technometrics* **1983**, *25*, 165–172. [CrossRef]
50. Gilbert, R.O. *Statistical Methods for Environmental Pollution Monitoring*; Van Nostrand Reinhold Co.: New York, NY, USA, 1987; ISBN 978-0-442-23050-0.
51. Millard, S.P. Package for Environmental Statistics, Including US EPA Guidance. 2022. Available online: https://cloud.r-project.org/web/packages/EnvStats/index.html (accessed on 28 October 2022).
52. Ramsay, J.O.; Silverman, B.W. *Functional Data Analysis*, 2nd ed.; Springer Series in Statistics; Springer: New York, NY, USA, 2005; ISBN 978-0-387-40080-8.
53. Torres, J.M.; Nieto, P.J.G.; Alejano, L.; Reyes, A.N. Detection of Outliers in Gas Emissions from Urban Areas Using Functional Data Analysis. *J. Hazard. Mater.* **2011**, *186*, 144–149. [CrossRef]
54. Rousseeuw, P.J.; Ruts, I.; Tukey, J.W. The Bagplot: A Bivariate Boxplot. *Am. Stat.* **1999**, *53*, 382–387. [CrossRef]
55. Tukey, J.W. *Exploratory Data Analysis*; Addison-Wesley Series in Behavioral Science; Addison-Wesley Pub. Co.: Reading, MA, USA, 1977; ISBN 978-0-201-07616-5.
56. Hyndman, R.J.; Shahid Ullah, M. Robust Forecasting of Mortality and Fertility Rates: A Functional Data Approach. *Comput. Stat. Data Anal.* **2007**, *51*, 4942–4956. [CrossRef]
57. Filzmoser, P.; Maronna, R.; Werner, M. Outlier Identification in High Dimensions. *Comput. Stat. Data Anal.* **2008**, *52*, 1694–1711. [CrossRef]
58. Hall, P.; Müller, H.-G.; Wang, J.-L. Properties of Principal Component Methods for Functional and Longitudinal Data Analysis. *Ann. Stat.* **2006**, *34*, 1493–1517. [CrossRef]
59. Ruts, I.; Rousseeuw, P.J. Computing Depth Contours of Bivariate Point Clouds. *Comput. Stat. Data Anal.* **1996**, *23*, 153–168. [CrossRef]
60. Hyndman, R.J. Computing and Graphing Highest Density Regions. *Am. Stat.* **1996**, *50*, 120. [CrossRef]
61. Scott, D.W. *Multivariate Density Estimation: Theory, Practice, and Visualization*, 2nd ed.; Wiley Series in Probability and Statistics; Wiley: New York, NY, USA, 2015; ISBN 978-0-471-69755-8.

62. Tanner, M.A. *Tools for Statistical Inference*, 2nd ed.; Springer Series in Statistics; Springer: New York, NY, USA, 1993; ISBN 978-1-4684-0194-3.
63. Saxena, P.; Sonwani, S. *Criteria Air Pollutants and Their Impact on Environmental Health*; Springer: Singapore, 2019; ISBN 9789811399916.
64. The European Parliament and the Council Parliament of the European Union. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on Ambient Air Quality and Cleaner Air for Europe; pp. 1–44. Available online: https://eur-lex.europa.eu/eli/dir/2008/50/2015-09-18 (accessed on 31 October 2022).
65. Government of the Principality of Asturias. Quality Air in Asturias 2021; 2022; p. 84. Available online: https://descargas.asturias.es/medioambiente/Estudios%20e%20informes%20calidad%20del%20aire/INFORME_CA_ASTURIAS_2021.pdf (accessed on 28 October 2022).
66. Ngarambe, J.; Joen, S.J.; Han, C.-H.; Yun, G.Y. Exploring the Relationship between Particulate Matter, CO, SO_2, NO_2, O_3 and Urban Heat Island in Seoul, Korea. *J. Hazard. Mater.* **2021**, *403*, 123615. [CrossRef]
67. Jury, M.R. Meteorology of Air Pollution in Los Angeles. *Atmos. Pollut. Res.* **2020**, *11*, 1226–1237. [CrossRef]
68. Han, X.; Naeher, L.P. A Review of Traffic-Related Air Pollution Exposure Assessment Studies in the Developing World. *Environ. Int.* **2006**, *32*, 106–120. [CrossRef] [PubMed]
69. Paraschiv, S.; Barbuta-Misu, N.; Paraschiv, S.L. Influence of NO_2, NO and Meteorological Conditions on the Tropospheric O_3 Concentration at an Industrial Station. *Energy Rep.* **2020**, *6*, 231–236. [CrossRef]
70. Gijon Port Authority Gijon Port Internal Report. Available online: https://www.puertogijon.es/musel-movio-mas-16-millones-toneladas-2020/ (accessed on 31 October 2022).
71. State-Owned Spanish Ports General Summary of Port Traffic. Traffic Statistics December 2021. Available online: https://www.puertos.es/es-es/estadisticas/Paginas/estadistica_mensual.aspx (accessed on 26 September 2022).
72. Moldanová, J.; Fridell, E.; Popovicheva, O.; Demirdjian, B.; Tishkova, V.; Faccinetto, A.; Focsa, C. Characterisation of Particulate Matter and Gaseous Emissions from a Large Ship Diesel Engine. *Atmos. Environ.* **2009**, *43*, 2632–2641. [CrossRef]
73. Deniz, C.; Kilic, A. Estimation and Assessment of Shipping Emissions in the Region of Ambarlı Port, Turkey. *Environ. Prog. Sustain. Energy* **2009**, *29*, 107–115. [CrossRef]
74. Deniz, C.; Kilic, A.; Cıvkaroglu, G. Estimation of Shipping Emissions in Candarli Gulf, Turkey. *Environ. Monit. Assess.* **2010**, *171*, 219–228. [CrossRef]
75. Mueller, D.; Uibel, S.; Takemura, M.; Klingelhoefer, D.; Groneberg, D.A. Ships, Ports and Particulate Air Pollution—An Analysis of Recent Studies. *J. Occup. Med. Toxicol.* **2011**, *6*, 31. [CrossRef]
76. Bartolomé Jimeno, M.; González López, L. Primer Informe de Evaluacion Conformidad Emplazamiento Estaciones Medicion Calidad Aire Principado Asturias—Zona Aglomeracion Gijon. 2021. Available online: https://www.isciii.es/QueHacemos/Servicios/SanidadAmbiental/SiteAssets/Paginas/LaboratorioNacionalReferenciaCalidadAire/Informe%20ISCIII%20de%20conformidad%20emplazamiento%20estaciones%20Asturias_Gij%C3%B3n.pdf (accessed on 27 October 2022).
77. Wang, Q.; Gu, J.; Wang, X. The Impact of Sahara Dust on Air Quality and Public Health in European Countries. *Atmos. Environ.* **2020**, *241*, 117771. [CrossRef]
78. Díaz Muñiz, C.; García Nieto, P.J.; Alonso Fernández, J.R.; Martínez Torres, J.; Taboada, J. Detection of Outliers in Water Quality Monitoring Samples Using Functional Data Analysis in San Esteban Estuary (Northern Spain). *Sci. Total Environ.* **2012**, *439*, 54–61. [CrossRef]
79. Chiou, J.-M.; Zhang, Y.-C.; Chen, W.-H.; Chang, C.-W. A Functional Data Approach to Missing Value Imputation and Outlier Detection for Traffic Flow Data. *Transp. B Transp. Dyn.* **2014**, *2*, 106–129. [CrossRef]
80. Hussain, I.; Uddin, M. Functional and Multivariate Hydrological Data Visualization and Outlier Detection of Sukkur Barrage. *IJCA* **2019**, *178*, 20–29. [CrossRef]
81. Ojo, O.T.; Fernández Anta, A.; Lillo, R.E.; Sguera, C. Detecting and Classifying Outliers in Big Functional Data. *Adv. Data Anal. Classif.* **2022**, *16*, 725–760. [CrossRef]