

Article

Robust Inferential Techniques Applied to the Analysis of the Tropospheric Ozone Concentration in an Urban Area

Wilmar Hernandez ^{1,*} , Alfredo Mendez ² , Vicente González-Posadas ³, José Luis Jiménez-Martín ³ and Iván Menes Camejo ⁴

- ¹ Facultad de Ingeniería y Ciencias Aplicadas, Universidad de Las Américas, Quito EC-170125, Ecuador
² Departamento de Matemática Aplicada a las Tecnologías de la Información y las Comunicaciones, ETS de Ingeniería y Sistemas de Telecomunicación, Universidad Politécnica de Madrid, 28031 Madrid, Spain; alfredo.mendez@upm.es
³ Departamento de Teoría de la Señal y Comunicaciones, ETS de Ingeniería y Sistemas de Telecomunicación, Universidad Politécnica de Madrid, 28031 Madrid, Spain; vicente.gonzalez@upm.es (V.G.-P.); joseluis.jimenez@upm.es (J.L.J.-M.)
⁴ Facultad de Informática y Electrónica, Escuela Superior Politécnica de Chimborazo, Riobamba EC-060155, Ecuador; imenes@esepoch.edu.ec
* Correspondence: wilmar.hernandez@udla.edu.ec

Abstract: This paper analyzes 12 years of tropospheric ozone (O₃) concentration measurements using robust techniques. The measurements were taken at an air quality monitoring station called Belisario, which is in Quito, Ecuador; the data collection time period was 1 January 2008 to 31 December 2019, and the measurements were carried out using photometric O₃ analyzers. Here, the measurement results were used to build variables that represented hours, days, months, and years, and were then classified and categorized. The index of air quality (IAQ) of the city was used to make the classifications, and robust and nonrobust confidence intervals were used to make the categorizations. Furthermore, robust analysis methods were compared with classical methods, nonparametric methods, and bootstrap-based methods. The results showed that the analysis using robust methods is better than the analysis using nonrobust methods, which are not immune to the influence of extreme observations. Using all of the aforementioned methods, confidence intervals were used to both establish and quantify differences between categories of the groups of variables under study. In addition, the central tendency and variability of the O₃ concentration at Belisario station were exhaustively analyzed, concluding that said concentration was stable for years, highly variable for months and hours, and slightly changing between the days of the week. Additionally, according to the criteria established by the IAQ, it was shown that in Quito, the O₃ concentration levels during the study period were not harmful to human health.

Keywords: robust analysis; central tendency and scale estimates; robust and nonrobust confidence intervals; categorization of ozone measurements



Citation: Hernandez, W.; Mendez, A.; González-Posadas, V.; Jiménez-Martín, J.L.; Camejo, I.M. Robust Inferential Techniques Applied to the Analysis of the Tropospheric Ozone Concentration in an Urban Area. *Sensors* **2021**, *21*, 277. <https://doi.org/10.3390/s21010277>

Received: 14 December 2020

Accepted: 30 December 2020

Published: 3 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Ozone is a substance that, based on its composition, is classified as a simple molecular substance made up of three oxygen atoms joined by covalent bonds, one single and the other double; its molecular formula is O₃ [1,2].

According to [3], the ozone molecule is angular, with an angle of 117° and a structure in resonance between two possible electronic configurations. It is diamagnetic, indicating the absence of unpaired electrons [3].

Ozone (O₃), whose name is trioxxygen according to [4,5], is one of the allotropes of oxygen, the most familiar of which is molecular oxygen (O₂), also known as dioxygen. Its CAS Registry Number (Chemical Abstracts Service registration number) [6] is 100-15-6.

1.1. Stratospheric Ozone

Ozone is concentrated in the stratosphere between 12 and 25 km high, forming the ozone layer, and reaching its maximum concentration at a height between 20 and 25 km. Stratospheric ozone is formed by the action of ultraviolet radiation, which dissociates the dioxygen molecules into two highly reactive atoms, as in Equations (1) and (2), which react with the molecular:



where O_2 is oxygen, λ stands for wavelength, the photon's energy is $E = h\nu$, the Planck constant is h , the photon's frequency is ν , and O is an oxygen atom. In addition, (2) occurs with heat release and requires the presence of a third body, M , such as N_2 (dinitrogen) or O_2 , which removes the energy from the reaction and stabilizes the ozone molecule.

Ultraviolet radiation with a wavelength greater than 290 nm causes the O_3 molecule to dissociate as shown in Equations (3) and (4):



This set of processes is called the Chapman cycle [7].

A dynamic equilibrium is formed in which ozone is formed and destroyed, consuming most of the wavelength radiation less than 290 nm. Thus, ozone acts as a filter that does not let harmful radiation pass through to the Earth's surface [8].

In accordance with [7], due to stratospheric winds, ozone is transported from tropical to polar regions. Thus, the closer human beings live to the equator, the less ozone will protect them from ultraviolet light.

The ozone balance in the stratosphere is affected by the presence of pollutants known as ozone-depleting substances, such as, among others, chlorofluorocarbons (CFCs), which rise to the upper atmosphere and catalyze the destruction of ozone faster than it is generated, thus producing a hole in the ozone layer [9].

1.2. Tropospheric Ozone

Tropospheric ozone, also called environmental ozone, is a colorless gas that is mainly created through photochemical reactions of nitrogen oxides (NO_x) and volatile organic compounds (VOCs) [10]. Additionally, soil and plants emit NO_x and VOCs (biogenic sources) [11].

O_3 is formed through reactions caused by sunlight. The phenomenon by which precursors are emitted and accumulated through reactions necessary for the formation of ozone to occur is called photochemical smog [12]. Smog episodes are the main production process for tropospheric ozone [13]. It is the most important oxidant present in the lower levels of the atmosphere, making it a potentially dangerous compound due to its ability to react with most compounds, degrading them. In addition, this affects both materials and living things that are affected by its action, externally but also internally, in the gas exchange that takes place when breathing, which is the main way humans are affected by ozone pollution. The response also varies greatly between individuals for genetic reasons (antioxidant capacity of cells), age (children and the elderly are the most sensitive groups), and the presence of respiratory conditions (allergies and asthma) [14].

When the ozone concentration is high, it is not advisable to do physical exercise, especially in the central hours of the day, since it increases the breathing rate and increases the entry of ozone into the lungs [11,14]. In regards to vegetation, a high level of ozone damages the leaves, reduces plant growth, and leads to a lower crop yield [11,14]. In the vicinity of the sources (cities, roads, and industries), fresh emissions can react with ozone

and locally reduce its concentration; nitric oxide emitted into the air locally by automobiles removes ozone carried from other parts [3]:



where NO is nitric oxide and NO₂ is nitrogen dioxide. However, at a certain distance, the photochemical formation of O₃ is reactivated, which is why the concentrations of O₃ are normally low in industrial centers and urban areas. On the other hand, in rural areas and in the outskirts of urban areas, the concentration is higher [11].

The photochemical formation of ozone does not occur during the night, because there is no sunlight [11]. In addition, during the day, the maximums of O₃ concentrations occur from noon, when the radiation is highest; regarding the time of year, the maximums occur during the spring and summer months.

The World Health Organization [15] has established that when the ozone concentration in the air that is breathed is higher than 240 µg/m³ and is maintained for more than 8 h, there is a probability of significant health effects [16].

To synthesize ozone, a machine called an ozone generator or ozonizer is used, in which a high electrical voltage known as the corona effect is generated, which produces ozone from O₂ [17]. The ozone generation has application in the elimination of bad odors and disinfection of the air, in the treatment and purification of water, and in ozone therapy.

What has been mentioned in the previous paragraphs provides a justification for carrying out a robust analysis [18–20] of the information given by ozone measuring systems, so that we can be able to describe with high precision the behavior of ozone concentration in cities, even when there are a large number of extreme observations and the data under study follow heavy-tailed distributions [19,21]. This last aspect has become a bottleneck because many times the outliers represent the response of the physical system to certain types of inputs. Therefore, these values are carriers of useful information and should not be eliminated, but must be treated appropriately in order to understand what they are telling us about the system from which they come. It is for this reason that the robust data analysis gains strength and shows a highly precise way to explain the behaviors of air pollution variables, even when they come from heavy-tailed distributions with a large number of outliers.

The aforementioned, applied to the analysis of ozone concentration in urban areas, constitutes one of the contributions of the research presented in this paper.

The robust analysis of the concentration of ozone at Belisario air quality monitoring station, which is in Quito, Ecuador, is the foremost objective of this research. This analysis is carried out from 1 January 2008 to 31 December 2019 [22,23].

Quito was chosen because of the effect of this city's vehicular traffic, poor fuel quality, and traffic jams on air pollution levels. Furthermore, there are other factors in the city that could increase air pollution levels, such as industrial zones and population growth [23].

The design of a sensor-based system aimed at measuring and georeferencing atmospheric variables, including ozone, is presented in [24]. In addition, a low-cost air quality monitoring station was presented in [25], where the authors proposed a calibration procedure based on artificial intelligence techniques. Furthermore, another air quality monitoring system was presented in [26], which employed a Zigbee network to improve the network layout. Additionally, a long short-term memory (LSTM) network was used in [26] to predict the urban air quality pollution period. Moreover, O₃ and seven other indicators were used in [26] to qualitatively validate the experiential knowledge of the authors; to allow cells to have long-term memory, the hidden layer of recurrent neural network cells was replaced by LSTM cells.

A meta-analysis of ground-level ozone pollution was carried out in [27]. This analysis was performed to find the similarities and differences between the risk of asthma aggravation and ground-level ozone exposure measurements. In [27], three different time periods (i.e., 1-h daily maximum, 8-h daily maximum, and 24-h average concentrations) were used. In addition, the meta-analysis performed in [27] was carried out using pooled relative risks, 95% confidence intervals, and random-effects models [28]. Additionally, in order

to achieve robustness in the results, subgroup and sensitivity analyses were conducted in [28]. Moreover, the Cochran Q statistic and I^2 estimation [29] were used to measure heterogeneity and inconsistency in the meta-analysis. Furthermore, in order to test whether there was publication bias, the authors made use of a funnel plot and Egger's test [30].

In [31], a study was carried out in 48 cities of China from 2013 to 2017. In that study, time-series analyses were conducted and the authors used a generalized model combined with a random effect model to estimate ozone levels. Another time-series study was performed in [32] in 184 cities of China from 2014 to 2017. Additionally, in [32], the relationship between patients due to pneumonia and ozone concentration was found by using a generalized additive model, and the authors provided robust proof of the existence of a relationship. The robustness of the relation was tested by fitting two-pollutant models.

A 10-year study on the relationship between particulate matter and ozone exposure (PM and OE), and a depression and anxiety diagnosis (DAD) in Saxony, Germany, was conducted in [33]. In that research, the data used for the analysis corresponded to the information collected from 2005 to 2014. In [33], the analysis was performed by using generalized estimating equations [34], and the robust metric was the number of days for which the maximum value of the 8-h average ozone concentration was greater than $120 \mu\text{g}/\text{m}^3$. Additionally, two-pollutant models were built and a sensitivity analysis was carried out, aimed at studying the relationship between PM and OE and DAD.

The reasons why there was a high ozone concentration in Chengdu, China in July 2017 were studied in [35]. In order to perform such a study, both measurements and air quality simulations were used. Additionally, in order to identify the VOC sources and perform the quantification of these sources, positive matrix factorization [36–38] was used. Moreover, in [35], the impact of physical and chemical processes on ozone concentrations was analyzed by the integrated process rate method [39–41].

In various situations in which it is intended to study the behavior of air pollution variables, researchers face the problem that these variables either do not follow a Gaussian distribution or do not follow any known parametric distribution. Therefore, classical statistical inference methods cannot be used for data analysis and nonparametric statistical inference must be used. For example, both Mann–Whitney U and Kruskal–Wallis tests [42, 43] were applied in [44] to analyze vehicle emissions. In [44], 1000 vehicles were tested in order to find significant differences between the mean emissions of air pollutants. Other instances of authors using robust techniques to analyze the concentration of air pollutants are presented in [45–54].

In this research, the measurement results of 12 years of tropospheric ozone concentration were analyzed using robust techniques. The urban area chosen for the study was Belisario station [22], and robust statistics [18–20] were used in this research to determine the robust central tendency and scale estimations of tropospheric ozone and find parametric, nonparametric, and robust confidence intervals that explain the O_3 concentration [18–20,42,43]. The O_3 concentration measurements that were analyzed in this paper were taken from 1 January 2008 to 31 December 2019.

The analysis presented here allowed us to judge the central tendency of the data based on the variability. Here, the data were grouped and classified, and similarities and differences were determined. Moreover, confidence intervals were used to do the aforementioned classification and several methodologies were used to analyze the data: classic, nonparametric, bootstrap, and robust methodologies.

In [23], another analysis of air pollution variables in Quito is presented. However, the analysis performed in [23] is not robust and, what is more, it is only based on the mean and maximum values. For this reason, the study performed in this research can be considered essential to comprehensively understand, in a formal and rigorous manner, the behavior of tropospheric ozone at Belisario monitoring station.

In this paper, in order to perform robust data analyses, each year under study was considered as a random variable. Additionally, it was shown that the distribution of these variables was heavy-tailed [19,21]. The concentration of other air pollution variables in

Quito was robustly analyzed in [47–51]. Furthermore, in [55,56] some robust estimators were also used. Additional research in which statistical tools have been used to analyze the O₃ concentration are as follows.

A pollution weather prediction system was proposed in [57] and used to measure O₃ among other pollutants. In [57], in order to carry out predictions, linear regression and artificial neural networks were used.

An air quality monitoring network aimed at analyzing the changeable nature of ozone across several communities of California, USA, was shown in [58], where the mean absolute error was used to analyze O₃ concentrations and the accuracy of measurement nodes and their correlation to reference instrumentation was indicated by using least squares regression. Moreover, summary statistics based on the mean, standard deviation, minimum, maximum, mean bias deviation, mean absolute deviations, and ordinary least squares statistics were used in [58] to present the data in a meaningful way.

In order to produce well-calibrated data, both multiple linear regression and nonlinear techniques were used in [59]. Additionally, a recalibration was done to mitigate the bias presented by the sensors and improve the variance.

A low-cost air quality monitoring system was presented in [60]. This system was aimed at monitoring O₃, among other pollutants, and a comparative analysis between a neuro-fuzzy system and a multilayer feed-forward perceptron was performed.

Finally, a statistical analysis of ensembles of O₃ profiles at the P. N. Lebedev Physical Institute, Moscow, Russia from 1996 to 2017 was carried out in [61]. This analysis was based on radiometric ozone monitoring and several statistical parameters were calculated: mean, variance, root-mean-square error, probability density function, probability distribution function, covariance function, correlation function, and frequency spectra.

The contribution of the present research with respect to the studies mentioned above is that, in order to optimize the sampling process to reduce power consumption in cases where researchers use portable devices powered by a battery, variables have been defined that represent the hours of the day in which the ozone concentration is the highest. This was done for both hours grouped by months and hours grouped by days of the week, separating working days from weekends, and it was shown that all the variables considered were different. For example, what happens in a particular month has nothing to do with what happens in other months. Therefore, it is difficult to make predictions since the distributions of the variables are different. Specifically, the variables do not come from the same statistical populations. Therefore, it is shown here that using robust methods is more effective than using nonrobust ones.

The objectives of this paper were as follows:

1. Compare the values of tropospheric ozone measurements based on four sets whose elements are variables that represent the following: (a) the 12 years under study, (b) the months: January to December, (c) the days: Monday to Sunday, and (d) the hours in pairs: from 0:00–1:00 to 22:00–23:00.
2. Analyze the behavior of the abovementioned variables in comparison with the different categories of air pollution established by the IAQ of Quito [23].
3. Estimate the data's central tendency and variability, and quantify differences using robust and nonrobust confidence intervals.

In this paper, it is shown that for the data under study, the trend of tropospheric ozone concentration at Belisario station has been towards stability for years, very variable by months and during the day, and moderately changing between the days of the week.

Some general comments on ozone sensors are made in Section 2. Section 3 describes the data and presents summary statistics on the collected data. Furthermore, nonparametric statistical inference is used in Section 4 to classify the data. Moreover, Section 5 is a robust data analysis and classification, and Section 6 is the discussion. Finally, Section 7 presents the conclusions of the paper.

2. Some General Comments on Ozone Sensors: Characteristics and Considerations of Signal Conditioning and Processing

In terms of low-cost O₃ measurements, the electrochemical (EC) sensor technology and the heated metal oxide sensor (HMOS) technology are at the vanguard [62]. Furthermore, the ultraviolet absorption principle-based analyzer is a well-known method for conducting O₃ measurements. However, the cost of O₃ analyzers is prohibitive [62].

In this research, O₃ measurements were carried out using THERMO 49C/49i photometric O₃ analyzers [63,64], which are used as measurement standards in several countries (e.g., these instruments are designated by the United States Environmental Protection Agency as a reference [65]). The principle of operation of the models 49C/49i is based on the fact that, at a wavelength of 254 nm, O₃ molecules absorb ultraviolet light.

The Beer–Lambert Law explains this type of absorption [63,64]. In short, a sample is introduced into the equipment and divided into two gas streams. After passing through an O₃ scrubber, one gas is used as the reference and sent to a solenoid valve. The other gas that is not sent to the O₃ scrubber is used as the sample gas and sent to another solenoid valve. Next, the reference gas and sample gas are deposited into two different cells and detectors are used to measure the ultraviolet light intensity of each cell. During the process of switching the sample gas, which is contained in one cell, and the reference gas, which is contained in the other cell, ultraviolet light intensity measurements are not performed until the two cells are flushed. Finally, both the O₃ concentration of each cell and the average concentration are calculated [63,64].

With regard to low-cost O₃ measurements, the principle of operation of EC and HMOS sensors is explained in [62]. Some of the strengths of EC sensors are as follows [62]: low power consumption, good repeatability and accuracy, and linear response, among others. On the other hand, some weaknesses of these sensors are that their readings can be affected by humidity and changes in temperature, they have a limited life, and their aging process is fast [62].

Furthermore, some of the strengths of HMOS sensors are as follows [62]: long life, very responsive to O₃ levels below 0.1 ppm, and excellent repeatability and accuracy. On the other hand, some weaknesses of these sensors are that their response time is slower than that of EC sensors, they require higher power consumption than EC sensors, and the linearity is reduced above 1 ppm [62].

In addition, when health and safety monitoring below 0.1 ppm is required, HMOS sensors can be used. However, when this monitoring is needed above 0.1 ppm, then EC sensors can be used. Moreover, EC sensors can be used to detect O₃ leaks, in controlling alarms and in O₃ generators (above 0.1 ppm). Additionally, HMOS sensors can be used in O₃ control applications below 0.1 ppm, when portable monitoring below 0.1 ppm is needed, and to measure ambient O₃ concentrations [62].

For the purpose of this research, in order to determine real-time O₃, a cumulative gas sensor is shown in [66]. Such a sensor used visible spectroscopy, which was based on cheap material, and was aimed at carrying out O₃ measurements in workplaces using a portable device.

At 600 nm, the abovementioned sensor showed high sensitivity at low O₃ concentrations, and good reproducibility and stability. In addition, a review of deep ultraviolet absorption applied to gas sensing is presented in [67], and a discussion of applications of ultraviolet absorption spectrometry for O₃ and other air pollutants is given in [67].

The performance of EC sensors is studied in [68], and it is shown that devices based on these sensors can be built and used at both lower cost and power consumption than conventional O₃ monitoring devices.

A review of gas and dissolved ozone sensors are presented in [69]. The different principles of operation of the sensors mentioned in [69] are based on the following: amperometric measurement methods, impedimetric measurement methods, and optical measurement methods. Moreover, this review was done in order to establish the requirements that these

sensors must meet in order to be used in medical applications. Ozone has also found a large number of applications in the search for solutions to SARS-CoV-2/COVID-19 [70,71].

Furthermore, another review of absorption spectroscopy-based O₃ sensors is given in [72]. In that paper, sensor applications, performance, limitations, and costs are discussed. Additionally, requirements such as accuracy, dynamic range, response, sensitivity, and cross-sensitivity are discussed for several specific applications.

However, it is not enough to design a sensor that measures the value of a certain physical quantity; this measurement must also be as robust as possible against disturbances and unwanted information. In this sense, several authors [66,67] are also concerned about the robustness that the sensors could have and are aware of the limitations of the sensor as an isolated element. Therefore, to achieve a good design of measurement equipment, it is necessary to put the sensor inside a signal conditioning and processing system that is capable of dealing with everything that represents noise, disturbance, and unwanted information [73]. Additionally, the stage of signal conditioning and processing must transform the sensor signal in such a way that it meets the design requirements of the measuring system.

It must be added that sometimes the systems that are designed to optimally solve the problem of conditioning and processing the signal from the sensors are not causal systems [74]. Therefore, it is necessary to resort to advanced design tools to make these systems realizable. However, the more one tries to design an optimal and robust system, the more complex and expensive the final design becomes. Therefore, it is recommended to establish a tradeoff between the desired performance of the signal conditioning and processing system and what can physically be achieved satisfactorily. Once this tradeoff is established, engineers and researchers can later use other analysis tools that allow them to robustly extract the relevant information from the signal under study. In general, in this last stage, the analysis of the information is carried out after the measuring system delivers the results to the user. Then, with the measurement results at hand, researchers and engineers focus on studying the characteristics of the data, the modeling, and the possibility of predicting the behavior of the physical quantity under study. In some cases, the aforementioned is done in order to design a robust feedback control system [75].

In this research, THERMO 49C/49i photometric O₃ analyzers were used to carry out the ozone measurements and, in order to extract the most important information from the measurement results, robust analysis methods were proposed. The proposed methods allowed us to classify and categorize variables that represented sets of ozone measurements, which were taken over hours, days, months, and years. Figure 1 shows the system diagram.

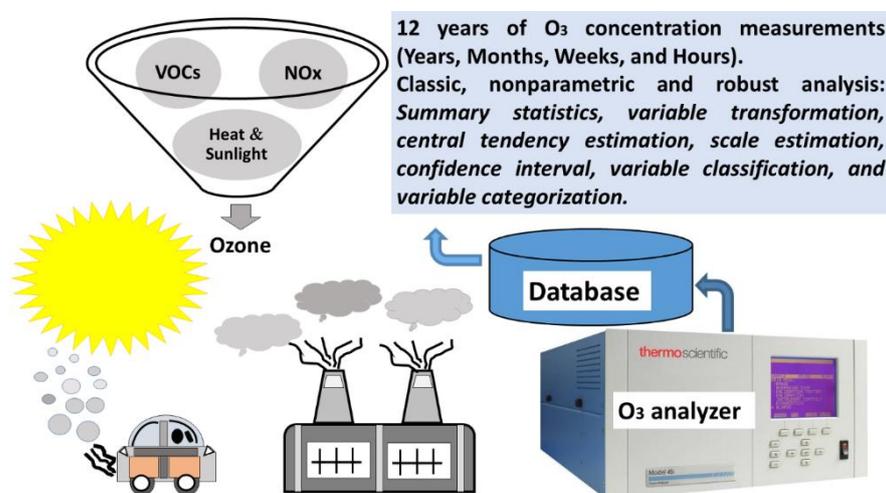


Figure 1. General system diagram.

The type of analysis of ozone measurements developed in this paper could serve as a starting point for design procedures that allow researchers to examine in detail the behavior of ozone in different parts of a city.

3. Data Description and Summary Statistics of Tropospheric Ozone Measurements at Belisario Station

Here, the ozone measurements were taken using THERMO 49C/49i photometric O₃ analyzers. The sampling rate was 1 h [23], which would imply that 105,193 data points were obtained. However, some observations could not be recorded because the data were not stored. The database was improved to take this circumstance into account, so it was possible to work with more than 96% of the data. In this way, it was guaranteed that more than 75% of the data were used for the analysis [76].

Due to the fact that the geographical and meteorological characteristics of Quito are adequate to have high insolation, the highest tropospheric ozone concentrations occur in August and September [23]. In addition, the months of the year that are either close to the equinox or correspond to the equinox are the abovementioned months, which are months in which the skies are clear. On the other hand, the months with the lowest tropospheric ozone concentrations in Quito are May and June, which are the months in which there are the most days with cloudy skies and rain [23].

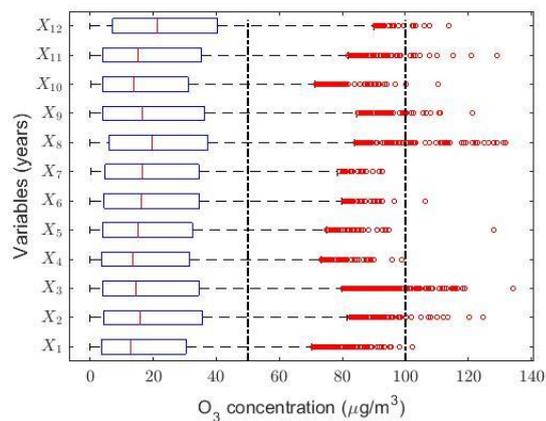
In order to find possible relationships between the data, the following variables were considered:

- $X_k, k = 1, \dots, 12$, where $X_1 = 2008, X_2 = 2009, \dots, X_{12} = 2019$.
- $Y_k, k = 1, \dots, 12$, where $Y_1 = \text{January}, Y_2 = \text{February}, \dots, Y_{12} = \text{December}$.
- $Z_k, k = 1, \dots, 7$, where $Z_1 = \text{Monday}, Z_2 = \text{Tuesday}, \dots, Z_7 = \text{Sunday}$.
- $W_k, k = 1, \dots, 12$, where $W_1 = 0 : 00-1 : 00, W_2 = 2 : 00-3 : 00, \dots, W_{12} = 22 : 00-23 : 00$.

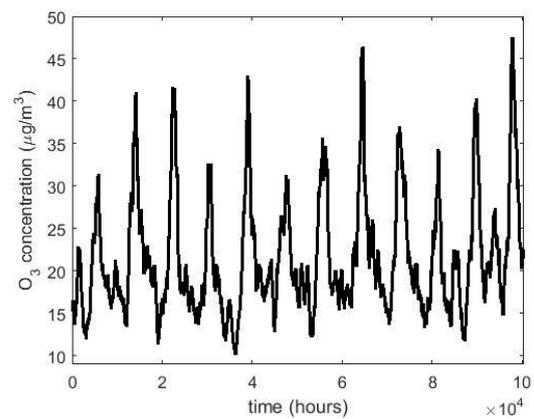
In order to summarize the observations, Table 1 shows some summary statistics. In addition, Figure 2a shows a multiple box plot of the variables $X_k, k = 1, \dots, 12$. In this figure, a separation between air pollution levels due to O₃ concentration in Quito [23] is made by using dashed vertical lines. These levels in $\mu\text{g}/\text{m}^3$ are as follows: Desirable = [0, 50); Acceptable = [50, 100); and Caution = [100, 200).

Table 1. Summary statistics: ozone observations.

Year (Variable)	Count	Mean of the Variable ($\mu\text{g}/\text{m}^3$)	Median of the Variable ($\mu\text{g}/\text{m}^3$)	Standard Deviation of the Variable ($\mu\text{g}/\text{m}^3$)	Skewness of the Variable	Kurtosis of the Variable	Minimum Value of the Variable ($\mu\text{g}/\text{m}^3$)	Maximum Value of the Variable ($\mu\text{g}/\text{m}^3$)	Outliers of the Variable (%)
2008 (X_1)	8741	19.2240	12.740	19.0235	1.1341	3.6247	0	102.12	1.60
2009 (X_2)	8371	22.1576	16.050	20.5857	0.9860	3.4136	0	124.54	0.79
2010 (X_3)	8447	21.8439	14.520	21.6142	1.2647	4.3375	0	134.08	2.12
2011 (X_4)	8392	19.3263	13.400	18.3904	0.9742	3.1587	0	98.89	0.54
2012 (X_5)	8484	20.1288	15.160	18.1590	0.8699	3.0665	0	128.08	0.40
2013 (X_6)	8277	21.0235	16.290	18.5324	0.7868	2.8418	0	106.31	0.28
2014 (X_7)	8506	21.3930	16.515	18.6472	0.7817	2.7286	0.13	92.57	0.24
2015 (X_8)	8509	23.6667	19.770	20.0291	0.9794	4.1337	0.04	131.68	0.92
2016 (X_9)	8384	22.6818	16.700	21.0013	0.9370	3.1916	0	121.22	0.63
2017 (X_{10})	8448	19.5031	13.875	17.9236	1.0155	3.4485	0.34	110.46	0.90
2018 (X_{11})	8405	21.6929	15.380	20.4016	0.9974	3.3810	0	129.17	0.76
2019 (X_{12})	8413	25.7776	21.470	20.7732	0.7496	2.8240	0	113.89	0.32
Total	101,107	21.5345	15.940	19.7131	0.9832	3.4724	0	134.08	0.80



(a) Box plots of the variables.



(b) Moving average of the time series that was built with all the observations.

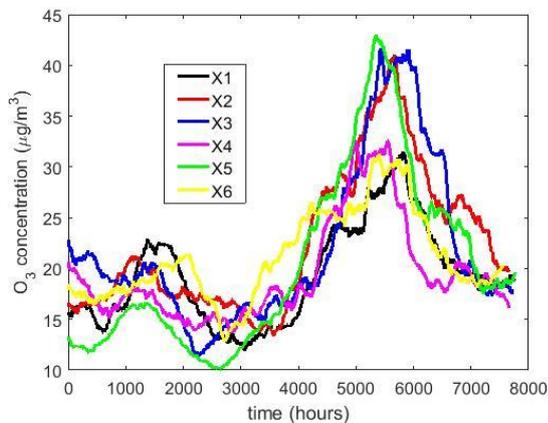
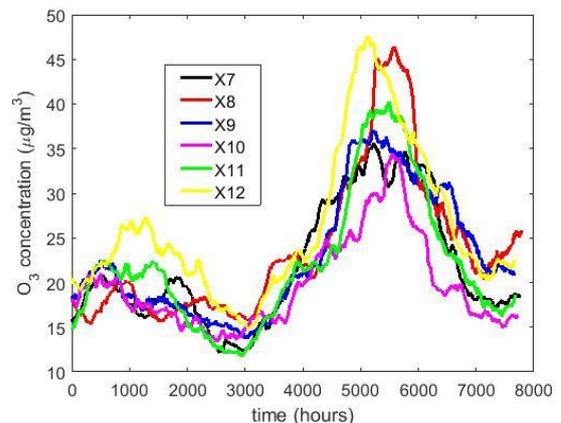
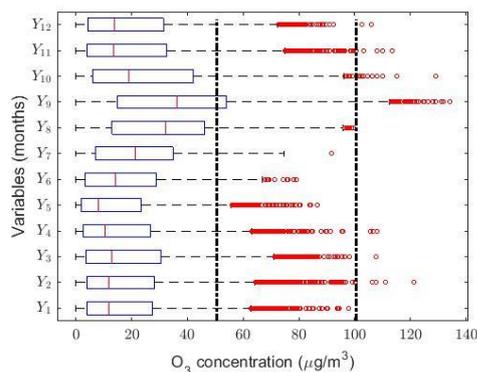
(c) Moving averages: $X_1 = 2008, \dots, X_6 = 2013$.(d) Moving averages: $X_7 = 2014, \dots, X_{12} = 2019$.

Figure 2. Multiple box plots of the variables (a) and moving averages (b–d) in the box plot diagrams, the red circles represent the outliers and, according to the IAQ of Quito [23], the vertical dashed lines indicate the following intervals of levels of air pollution in $\mu\text{g}/\text{m}^3$ due to O_3 concentration in Quito: Desirable = $[0, 50)$; Acceptable = $[50, 100)$; and Caution = $[100, 200)$.

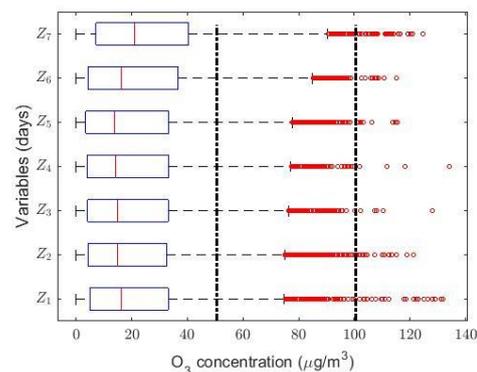
In all the variables shown in Figure 2a, the following comparisons can be made: mean $>$ median, skewness $>$ 0, and kurtosis $>$ 2.7. Furthermore, kurtosis $>$ 4 in both 2010 and 2015. All of the aforementioned indicates that the distributions of X_k , $k = 1, \dots, 12$, are heavy-tailed [19,21], or that these variables can be considered due to the existence of a mixture of distributions. In addition, in Figure 2a it can be seen that each year presents observations above the desirable level. Additionally, it is observed that abnormally high observations are present almost every year, although the percentage of such observations does not exceed 2.2%. Moreover, it is observed that the desirable level is always far exceeded, although the average does not exceed these values by much. All this confirms, once again, that the samples do not come from a normal distribution.

The moving averages (MAs) [77,78] of all years are shown in Figure 2b, the MAs of the years from 2008 to 2013 are shown in Figure 2c, and the MAs of the years from 2014 to 2019 are shown in Figure 2d. In this research, the size of the moving average MA was equal to 720, which represents the number of samples that can be taken in a month, assuming that each month has a duration of 30 days. The graph of MAs of all years indicates a stable concentration of O_3 . Furthermore, the maximum and minimum are reached in the third and second quarters, respectively. This corroborates the findings of [23]. Additionally, it is observed that when the observations are smoothed, none of the values fall outside the interval of desirable air pollution values. Therefore, the overcoming of this level is in specific moments and, when this occurs, it does not take place in a sustainable way.

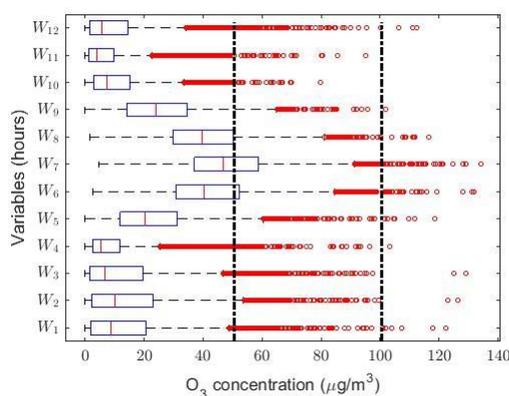
The analysis performed by months, days, and hours confirmed the previous conclusions. Figure 3a–c shows that the observations have heavy-tailed distributions. In addition, Figure 3a shows that the O₃ concentration is lowest in the summer months, although the maximums occur in September and October. There do not appear to be any differences in O₃ concentration when the analysis is carried out by days (see Figure 3b). Furthermore, Figure 3c shows that the O₃ concentration grows slightly at noon and that there are a lot of abnormally high observations in all time slots.



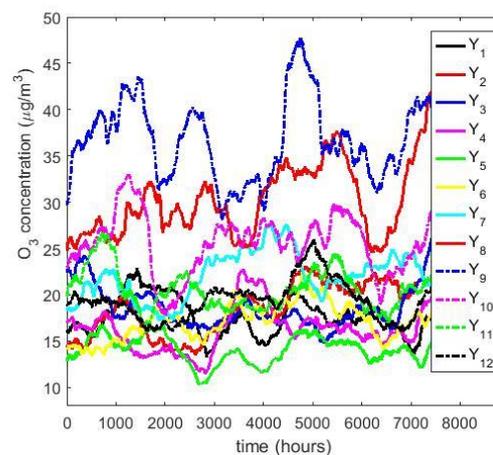
(a) Box plot: Months; Y₁ = January, ..., Y₁₂ = December.



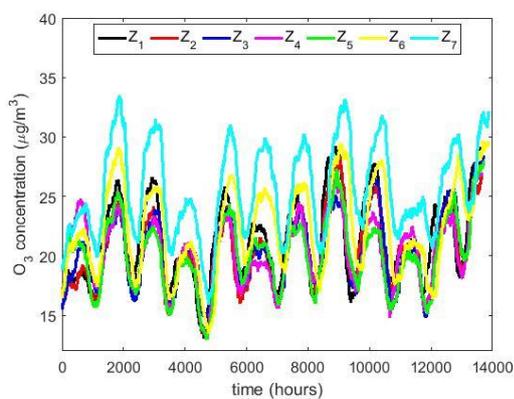
(b) Box plot: Days; Z₁ = Monday, ..., Z₇ = Sunday.



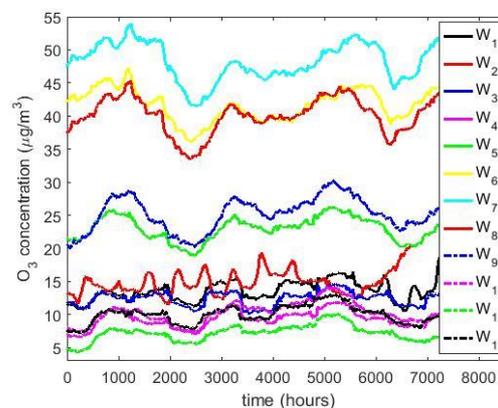
(c) Box plot: Hours; W₁ = 0:00 – 1:00, ..., W₁₂ = 22:00 – 23:00.



(d) Moving averages: Months; Y₁ = January, ..., Y₁₂ = December.



(e) Moving averages: Days; Z₁ = Monday, ..., Z₇ = Sunday.



(f) Moving averages: Hours; W₁ = 0:00 – 1:00, ..., W₁₂ = 22:00 – 23:00.

Figure 3. Box plots of the variables (a–c) and moving averages (d–f). In the box plot diagrams, the red circles represent the outliers and, according to the IAQ of Quito [23], the vertical dashed lines indicate the following intervals of levels of air pollution in μg/m³ due to O₃ concentration in Quito: Desirable = [0, 50); Acceptable = [50, 100); and Caution = [100, 200).

Figure 3d–f shows that the months, days, and hours have similar behavior throughout the years. In short, if the O₃ concentration values are higher or lower in any period of a particular year or day, something similar occurs for the rest of the years. In the case of the analysis by days (see Figure 3e), more noticeable periodic behavior appears. Moreover, it is possible that the periodic behavior observed in Figure 3e also manifests itself in Figure 3d,f. Therefore, it is likely that the samples come from a variable that presents behavior patterns that repeat every certain time interval. However, the frequencies of the possible oscillations that can be seen in Figure 3d–f seem to be neither the same nor integer multiples of the others. Additionally, it does not seem that some frequencies are rational multiples of others. Therefore, it can be said that this type of signal could have characteristics that are typical of variable frequency signals, among other things. The mathematical modeling of this type of signal belongs to different research that is not part of the objective of this paper.

In this research, different transformations of variables were made [79], because classical statistical inference methods could not be used. The previous paragraphs have shown that the data distribution was not normal. Nevertheless, this strategy did not give satisfactory results, because the data of the years could only be fitted to heavy-tailed distributions.

Furthermore, in order to analyze the ozone concentration by months, weeks, and hours, several transformations of variables were also attempted, and some transformations of a very few variables produced adjustments close to the logistic or normal distributions. However, when these changes were attempted for the rest of the variables of the same type, it was not possible to achieve distributions that were not heavy-tailed.

Therefore, the aforementioned was the motivation for using both nonparametric statistics and robust statistics to perform the type of data analysis presented in this paper.

4. Nonparametric Analysis

In this paper, after having found that there was almost no linear correlation between samples, it was concluded that they came from linearly independent variables. In addition, the samples were compared with each other by making use of the Wilcoxon signed-rank test (WSRT) [43]. Here, $M_e = M_0$ was the null hypothesis (H_0), where M_e stands for the median; the alternative hypothesis (H_1) was $M_e \neq M_0$. Specifically, if H_0 is true and observations behave in a stable manner, then half of the observed values are less than M_0 .

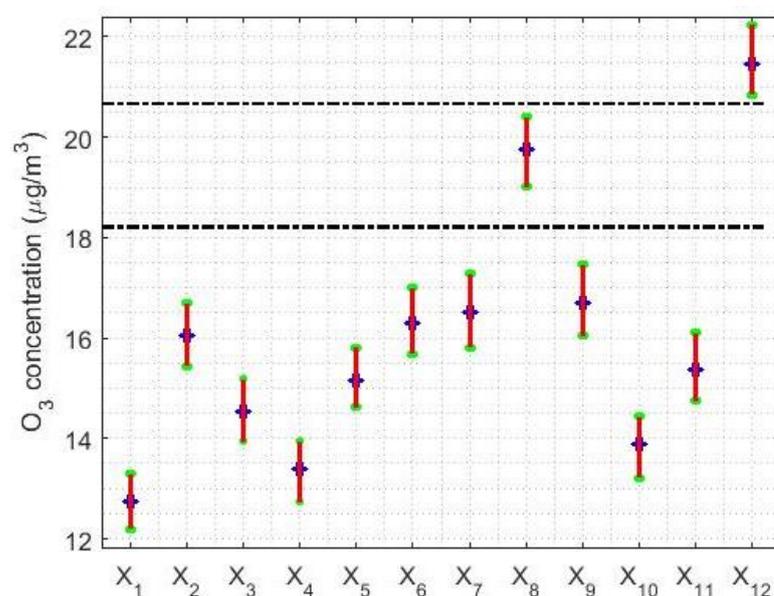
In the case under study, at the $(1 - \alpha)$ confidence level [43], the nonparametric confidence intervals should be as follows: $P\left(X_{(k'_{\alpha/2})} < M_e < X_{(k_{\alpha/2})}\right) = 1 - \alpha$, where $k'_{\alpha/2} = N/2 + 0.5 - z_{\alpha/2}\sqrt{N/4}$, $k_{\alpha/2} = N/2 + 0.5 + z_{\alpha/2}\sqrt{N/4}$, N is the sample of length, and, at $(1 - \alpha)/2$, $z_{\alpha/2}$, depends on the normal distribution [43,79].

The limits of the confidence intervals are shown in Table 2, and Figure 4 shows a graphical representation of these intervals. From the analysis of Table 2 and Figure 4, we see that no trend in the concentration of O₃ is clear. It can be seen that there were three years, 2008, 2011, and 2017, that seemed to have lower O₃ concentrations. However, there is an overlap of intervals with all the years, with two exceptions: 2015 and 2019. These two years have higher O₃ concentration levels than the rest of the years, especially 2019. In addition, 2015 and 2019 are different from each other. On the other hand, the widths of the confidence intervals have no significant differences. Moreover, it is verified once again that the desirable level of air pollution [23] was exceeded.

Furthermore, Figure 4 shows that the data can be grouped into three categories. One formed by the variables X_8 (corresponding to 2015), another formed by the variable X_{12} (corresponding to 2019), and the third formed by the rest of the years. The WSRT [43] showed that the variables' median homogeneity was the same as that obtained using nonparametric confidence intervals, although with some small differences because p -values [43] greater than 5% were considered. For example, X_1 is significantly different from the distribution of X_3 , but not different from X_4 . Moreover, X_3 is also not significantly different from X_4 .

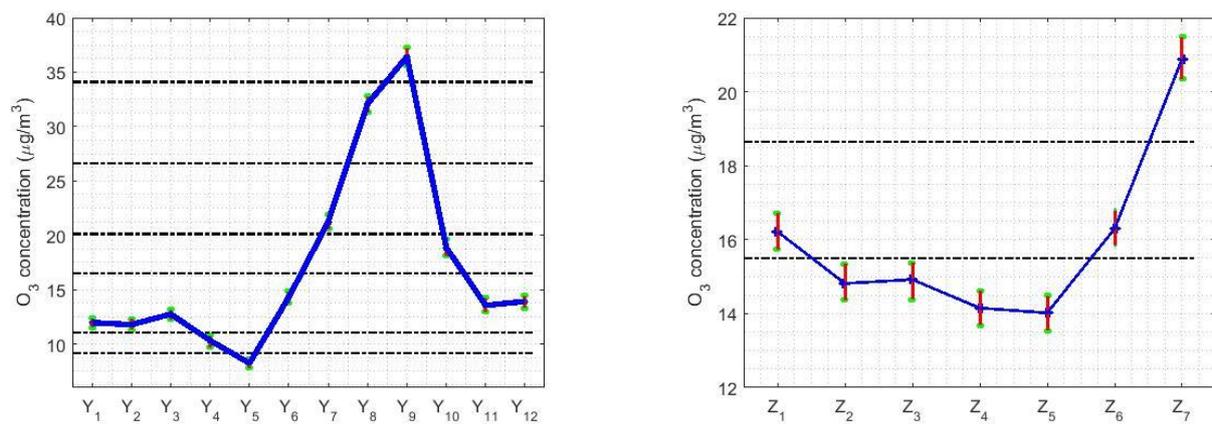
Table 2. The 95% confidence intervals ($CI_{0.95}$): median.

Variable (Year)	Lower Limit of the $CI_{0.95}$ ($\mu\text{g}/\text{m}^3$)	Upper Limit of the $CI_{0.95}$ ($\mu\text{g}/\text{m}^3$)
X_1 (2008)	12.20	13.29
X_2 (2009)	15.44	16.69
X_3 (2010)	13.95	15.18
X_4 (2011)	12.74	13.95
X_5 (2012)	14.62	15.81
X_6 (2013)	15.69	17.00
X_7 (2014)	15.81	17.28
X_8 (2015)	19.02	20.41
X_9 (2016)	16.05	17.47
X_{10} (2017)	13.22	14.44
X_{11} (2018)	14.75	16.10
X_{12} (2019)	20.85	22.24

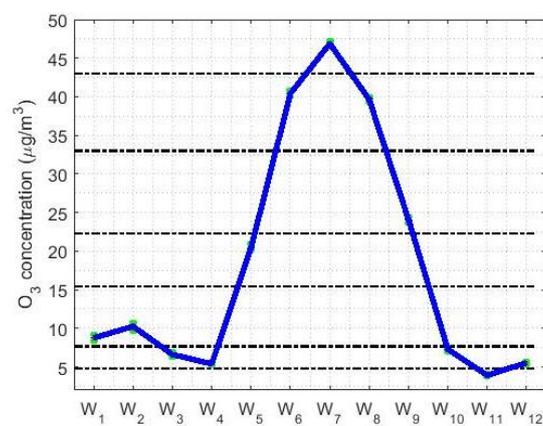
**Figure 4.** The 95% confidence intervals. Nonparametric intervals for the median: $X_1 = 2008, \dots, X_{12} = 2019$. The separation of categories is carried out using horizontal dashed lines.

For the nonparametric analysis of the months (see Figure 5a), seven categories were established. In this case, the use of the nonparametric intervals and the WSRT yielded the same results. According to the months, the level of O_3 concentration seems to be a periodic signal. At the end of summer, the ozone concentration levels are the highest, when there is less activity in the city, and these levels reach their minimum value in May. In addition, a stabilization occurs between the first and last months of the year, and between May and November a rebound occurs and the stabilization level is reached. Additionally, the nonparametric confidence intervals' amplitude was directly proportional to the value of the median.

Figure 5b shows that, when analyzing the weeks, the concentration of O_3 increased notably on the weekends and continued to decrease on weekdays. In this case, the WSRT and the nonparametric intervals yielded the same results. Furthermore, the width of the nonparametric confidence intervals was directly proportional to the value of the median.



(a) Confidence intervals: $Y_1 = \text{January}, \dots, Y_{12} = \text{December}$. (b) Confidence intervals: $Z_1 = \text{Monday}, \dots, Z_7 = \text{Sunday}$.



(c) Confidence intervals: $W_1 = 0:00\text{--}1:00, \dots, W_{12} = 22:00\text{--}23:00$.

Figure 5. The 95% nonparametric confidence intervals for the median. The separation of categories is carried out using horizontal dashed lines.

Finally, in the analysis of hours (see Figure 5c), the level of O₃ concentration reached its maximum value, close to 50 μg/m³, around 12:00. For the rest of the hours, the concentrations were close to 5 μg/m³ around 7:00 and around 22:00. Moreover, there were more categories than for the analysis of the days of the week, because between states of high concentration and low concentration there were transition variables. Between 22:00 on the current day and 7:00 on the following day, the O₃ concentration remained stable, with minor changes.

5. Robust Analysis

The analysis carried out in this section is based on robust statistics, which are not sensitive to extreme observations. As a consequence, the estimations of the data distribution and dispersion performed in this section had little sensitivity to the influence of extremely high or low observations [18–20]. In addition, the foremost objective of this section is to analyze the observations by using these estimations.

Furthermore, in order to prevent robust estimators from being affected by extreme observations, the influence curve [80] is used to characterize the robust statistics of this paper. Moreover, the order sample statistics [43] is used to obtain the robust estimators found in the paper.

5.1. Location and Scale Estimators

Location estimators [18–20,81]. TM : Trimean; $T_{wa}(c)$: Andrew's wave; $T(\alpha)$: α -trimmed mean; $T_{bi}(c)$: Biweight; and $W(\alpha)$: α -winsorized mean.

Scale estimators [18–20,80,82,83]. s_x : Sample standard deviation; MAD : Median absolute deviation; MAD_{mean} : Mean absolute deviation; $s^W(\alpha)$: Winsorized standard error; SRH : One-half of the fourth-spread; $s_{\omega a}(c)$: Andrew’s wave; C_n^α : Estimator based on a subrange; LMS : Least median squares; and $S_{bi}(c)$: Biweight.

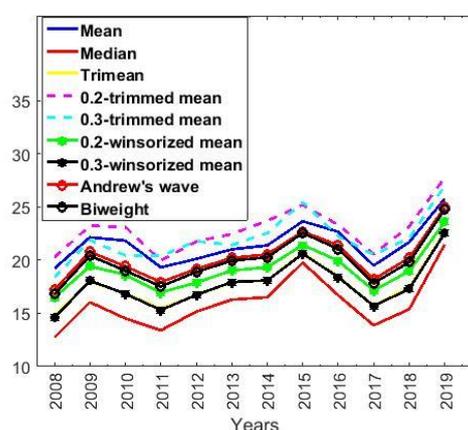
Tables 3 and 4 show the point estimates of location and scale, respectively. In addition, the above estimators are graphically represented in Figure 6.

Table 3. Estimates of the location of the observations: point estimates ($\mu\text{g}/\text{m}^3$).

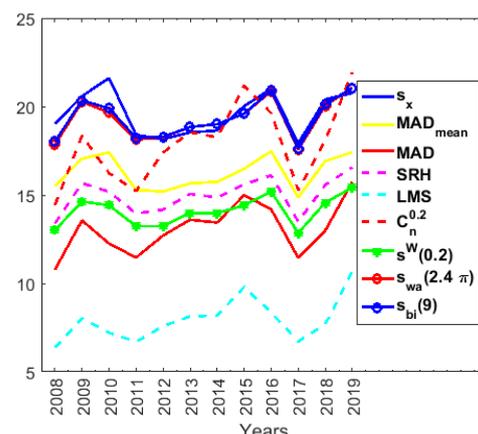
Year (Variable)	Mean	M_e	TM	$T(0.2)$	$T(0.3)$	$W(0.2)$	$W(0.3)$	$T_{wa}(2.4\pi)$	$T_{bi}(9)$
2008 (X_1)	19.2240	12.7400	14.8475	20.2708	18.3577	16.4755	14.5822	17.2095	16.7942
2009 (X_2)	22.1576	16.0500	17.9700	23.2451	21.8878	19.4422	18.0712	20.7875	20.4229
2010 (X_3)	21.8439	14.5200	16.9025	23.1257	20.4529	18.6331	16.8467	19.4255	19.0174
2011 (X_4)	19.3263	13.4000	15.4825	19.9706	20.4038	16.9701	15.3045	17.9404	17.5358
2012 (X_5)	20.1288	15.1600	16.6900	21.7689	21.8030	17.8895	16.7320	19.1783	18.8911
2013 (X_6)	21.0235	16.2900	17.9100	22.4651	21.4295	19.0371	17.9505	20.2210	19.9591
2014 (X_7)	21.3930	16.5150	18.0950	23.7089	22.5498	19.3308	18.0997	20.5514	20.2711
2015 (X_8)	23.6667	19.7700	20.7250	25.1945	25.4667	21.4437	20.6082	22.7003	22.5663
2016 (X_9)	22.6818	16.7000	18.4675	23.3377	22.2582	19.9581	18.4120	21.4043	21.0557
2017 (X_{10})	19.5031	13.8750	15.7275	20.5899	20.5212	17.1472	15.6978	18.1965	17.8385
2018 (X_{11})	21.6929	15.3800	17.4650	23.1852	22.0576	18.9715	17.2909	20.2167	19.8159
2019 (X_{12})	25.7776	21.4700	22.6300	27.7630	26.9957	23.5995	22.5443	24.9686	24.7720
All years	21.5345	15.9400	17.7425	21.3231	21.3523	19.0408	17.6475	20.2792	19.9625

Table 4. Estimates of the scale of the observations: point estimates ($\mu\text{g}/\text{m}^3$).

Year (Variable)	s_x	MAD_{mean}	MAD	SRH	LMS	$s^W(0.2)$	$s_{wa}(2.4\pi)$	$s_{bi}(9)$	$C_n^{0.2}$
2008 (X_1)	19.0235	15.5054	10.780	13.385	6.370	13.0238	17.8466	18.0195	14.4307
2009 (X_2)	20.5857	17.0561	13.560	15.670	8.025	14.6313	20.2597	20.3431	18.3482
2010 (X_3)	21.6142	17.4023	12.260	15.185	7.220	14.4179	19.6660	19.9031	16.2346
2011 (X_4)	18.3904	15.2864	11.460	13.955	6.700	13.2390	18.1400	18.2282	15.2142
2012 (X_5)	18.1590	15.1994	12.710	14.170	7.580	13.2382	18.2522	18.2864	17.4007
2013 (X_6)	18.5324	15.6567	13.600	15.070	8.145	13.9860	18.8507	18.8497	18.5486
2014 (X_7)	18.6472	15.7440	13.445	14.855	8.185	13.9803	19.0047	19.0099	18.2571
2015 (X_8)	20.0291	16.4902	14.990	15.600	9.805	14.4142	19.5946	19.5970	21.1724
2016 (X_9)	21.0013	17.4754	14.190	16.105	8.350	15.1890	20.8681	20.9305	19.6236
2017 (X_{10})	17.9236	14.8651	11.445	13.500	6.720	12.8552	17.5676	17.6466	15.2871
2018 (X_{11})	20.4016	16.9178	12.990	15.570	7.690	14.5733	20.0564	20.1505	18.0931
2019 (X_{12})	20.7732	17.4184	15.750	16.570	10.695	15.4182	21.0559	21.0423	21.9559
All years	19.7131	16.3365	13.210	15.095	7.970	14.1425	19.4223	19.4884	17.9655



(a) Estimates of location: 2008–2019.



(b) Estimates of scale: 2008–2019.

Figure 6. Estimates of location and scale by year.

In Figure 6a, according to the location estimates, the ozone concentration is stable, with two increases in 2015 and 2019; the rise occurring in 2019 is more pronounced. It is interesting to see that the 0.2-trimmed mean and the median are the boundaries of the location measures.

In Figure 6b, according to the scale estimates, the standard deviation and the *LMS* point estimator are the boundaries of the scale measures, with the biweight midvariance being very similar to the Andrew's wave.

Likewise, it is noteworthy that the scale estimates are quite stable for each estimator and very large compared to the location estimates. These estimates indicate great variability in the O_3 concentration measurements, which, on the one hand, have few outliers, but, on the other, have many high observations compared to the center of the distribution. What was said previously was already appreciated when analyzing Figure 2a.

So that the O_3 concentration analysis could also be carried out over periods of time that are representative of the daily activity of human beings, we decided to repeat what was done previously for the analysis of the years, but this time to analyze the O_3 concentration by months, days, and hours. Figure 7 shows graphical representations of the new estimates.

The location estimates of the variables of the months (see Figure 7a) are all very similar to each other; the higher the value of the estimate, the more similar these estimates are. Additionally, the scale estimates are very stable (see Figure 7d). The boundaries of these estimates are s_X and $C_n^{0.2}$ (upper boundary) and the *LMS* point estimator (lower boundary). Furthermore, despite having few outliers, the variability is high because the value of the scale estimates cannot be discarded.

Figure 7b shows that the O_3 concentrations increase on the weekends, keeping the location estimates stationary. The O_3 concentration during the weekend is higher than during the first days of the week. All estimates are bounded by the median and the mean.

Figure 7e shows that the scale estimates evolve in parallel in a band of values. It is notable that the lower bound of all these estimates are the estimates given by the *LMS* point estimator, just as it happens in Figure 7d,f. In other words, the *LMS* scale estimates are much lower than the other estimates. Furthermore, the biweight, standard deviation, and Andrew's wave values are very high. This result suggests that there is high variability without the need for extreme observations to be all outliers because the biweight and the Andrew's wave are robust.

In the day location estimates (see Figure 7c), from 7:00 there is a very abrupt rise in the concentration of O_3 until noon, going from $5 \mu\text{g}/\text{m}^3$ to about $48 \mu\text{g}/\text{m}^3$. Afterward, there is also a very pronounced drop until 20:00, which is when the commercial activity ends. For the rest of the hours, the O_3 concentration is similar, remaining at very low levels.

Regarding the scale estimates of the hours of the day (see Figure 7f), these estimates have a behavior similar to that of the other variables. In short, the boundaries of these estimates are the standard deviation and the *LMS* estimator. In general, these scale estimates appear to differ from each other in a constant and the increase is related to the increase in the location estimate.

Finally, it is worth mentioning that the same pattern seems to be followed by the months, weeks, and hours. We observed that, for each variable, the curves grew and decreased at the same time. Furthermore, we observed that the more the ozone concentration increased, the more the variability increased, and vice versa.

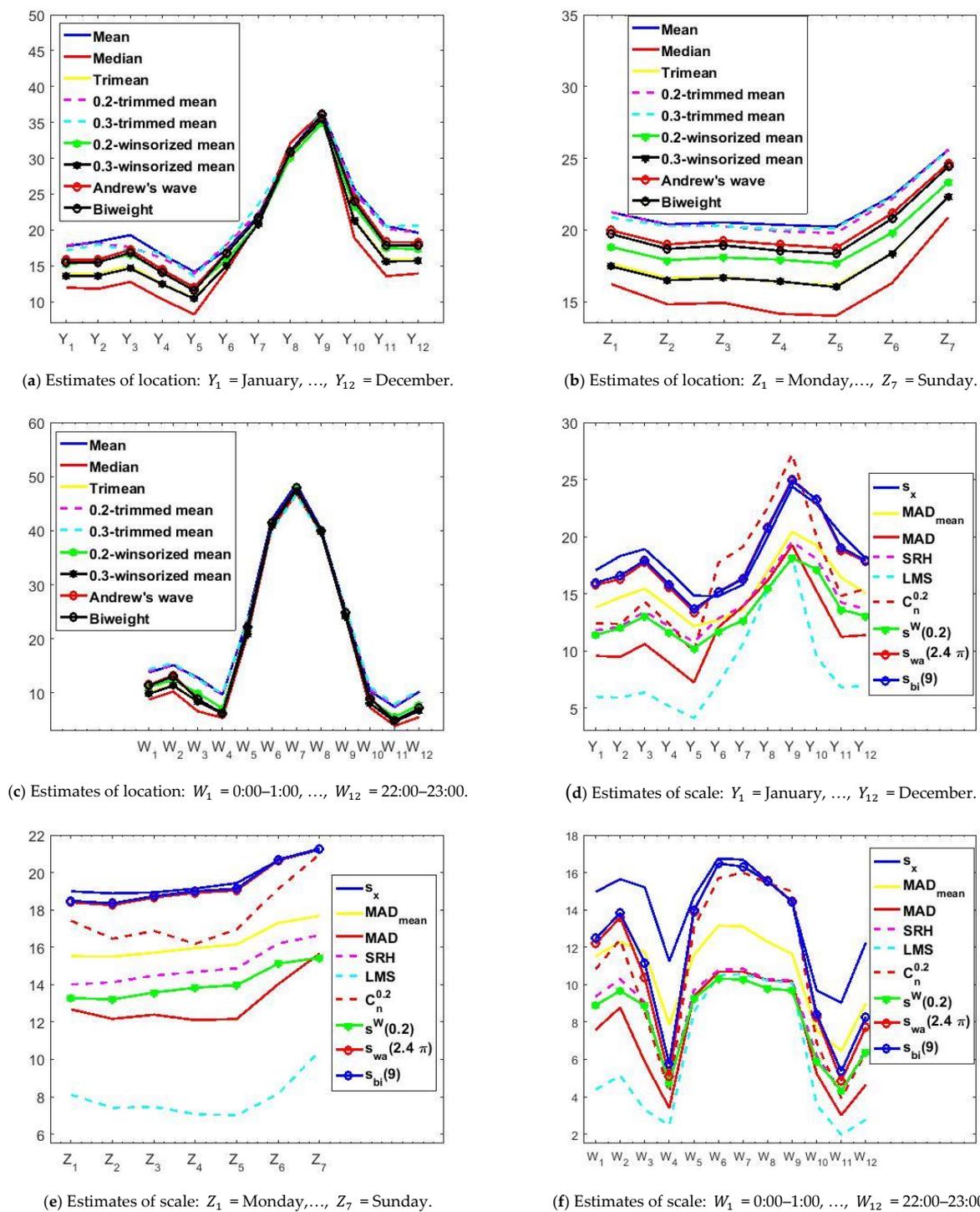


Figure 7. Estimates of location and scale: month, weeks, and hours.

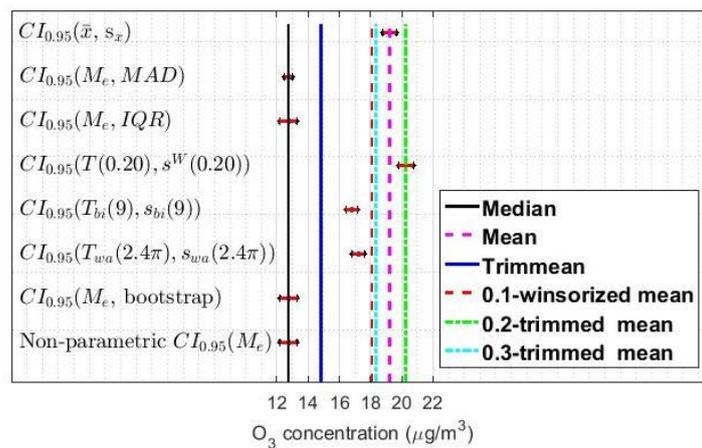
5.2. Confidence Intervals

In this research, the confidence intervals were built using the methodology explained in [48–50], which is in accordance with [84,85]. These intervals allowed us to perform the variable classification and categorization and recognize similarities and differences. The intervals are as follows:

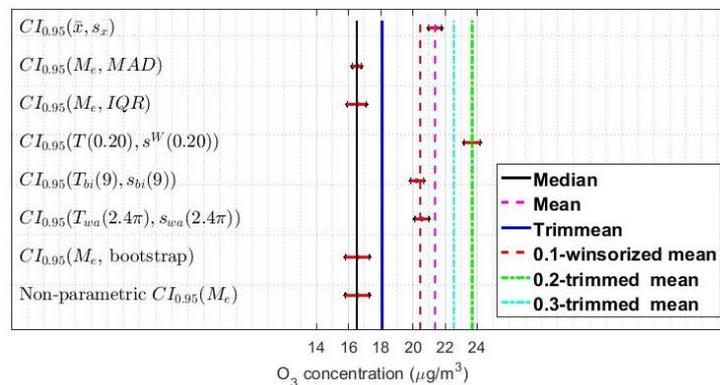
- (\bar{X}, s_x) , where \bar{X} is the mean
- (M_e, MAD) , where M_e is the median
- (M_e, IQR) , where IQR is the interquartile range

- $(T(\alpha), s^W(\alpha))$
- $(T_{wa}(c), s_{wa}(c))$
- $(T_{bi}(c), s_{bi}(c))$.

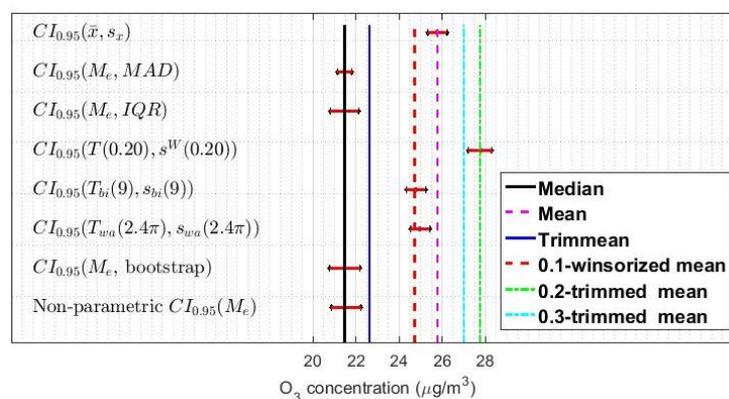
In addition, other confidence intervals were built by using a bootstrap method [20,49,50]. As a result, eight confidence intervals were built: one classic, one bootstrap, one parametric, and five robust. The intervals have been included in this paper for three of the 12 variables under study. Specifically, intervals are shown here for 2008, 2014, and 2019 (see Figure 8). That is, the intervals are shown for the two end variables and one intermediate. Showing these intervals for more variables would not contribute significantly to the analysis carried out in this section.



(a) 95% confidence intervals: 2008 (X_1).



(b) 95% confidence intervals: 2014 (X_7).



(c) 95% confidence intervals: 2019 (X_{12}).

Figure 8. The 95% confidence intervals ($CI_{0.95}$): 2008 (X_1), 2014 (X_7), and 2019 (X_{12}).

Analyzing Figure 8, it can be concluded that all the variables have similar characteristics. First, the 0.2-trimmed mean and winsorized variance-based confidence intervals are the most shifted to high concentrations, closely followed by the classic confidence intervals. The difference is approximately $2 \mu\text{g}/\text{m}^3$. Second, the biweight confidence interval is very similar to the Andrew's wave confidence interval. Third, the intervals with the lowest values are those based on the median. Furthermore, a similarity between nonparametric and bootstrap intervals is observed. Additionally, the pair (M_e, MAD) yields the narrowest confidence intervals. Finally, it should be noted that there is a slight increase in estimates for the early years compared to later years.

According to what has been previously said, the estimators $(T(\alpha), s^W(\alpha))$ and $(T_{bi}(c), s_{bi}(c))$ were used in this paper to compare the variables under study, for the following reasons:

1. Classic confidence intervals are based on the fact that an approximately normal underlying distribution is assumed. However, in this research that is not the case.
2. The (M_e, MAD) and (M_e, IQR) point estimators, the bootstrap estimators, and the nonparametric estimators give similar results. Nevertheless, when groupings of variables are analyzed, the pair (M_e, MAD) allows for establishing a greater number of differences between them.
3. The estimators based on the biweight and Andrew's wave yield similar results. Therefore, it is acceptable to use one type of estimator or the other to carry out the comparative analysis.

Information on the confidence intervals for the pair of estimators $(T(0.2), s^W(0.2))$ and $(T_{bi}(9), s_{bi}(9))$ is given in Table 5. In addition, graphical representations of these intervals are shown in Figure 9. Moreover, horizontal lines have been used in Figure 9 to carry out the classification of the variables, which is the same as that performed in Section 4. The only difference between the classification obtained in this section and the classification obtained using nonparametric estimators (see Section 4), is that the variables $X_1, X_4,$ and X_{10} can be grouped in the same category for the pair of estimators $(T(0.2), s^W(0.2))$ and for the pair of estimators $(T_{bi}(9), s_{bi}(9))$ in a clearer way.

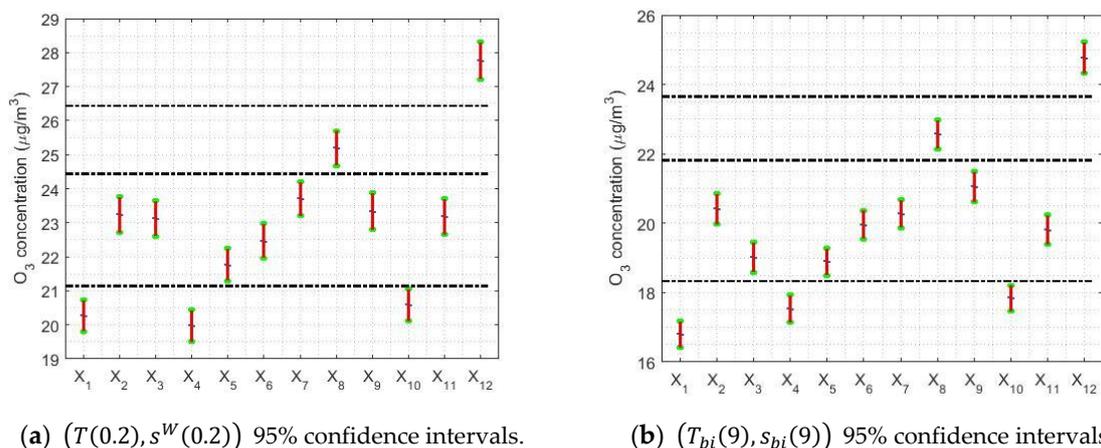


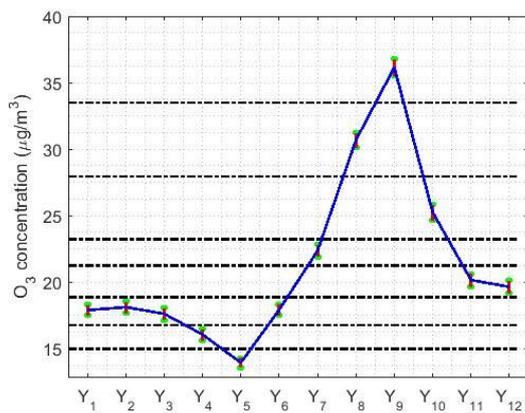
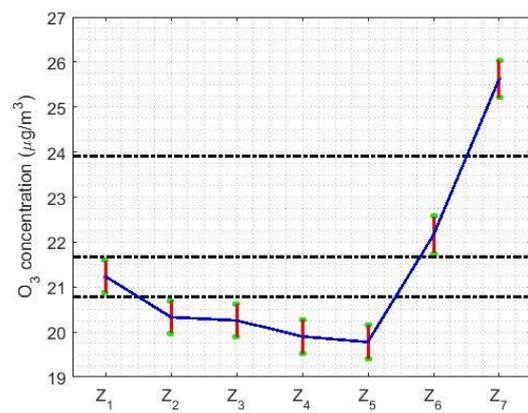
Figure 9. The 95% confidence intervals: $X_1 = 2008, X_2 = 2009, \dots, X_{12} = 2019$. The separation between categories is made by the horizontal dashed lines.

Regarding the amplitudes of these confidence intervals, it can be said that the pair $(T(0.2), s^W(0.2))$ yielded confidence intervals that are approximately 15% wider than those obtained with the pair $(T_{bi}(9), s_{bi}(9))$. In addition, the median and amplitude of these intervals evolve in parallel.

Figure 10 shows graphs that are used to study confidence intervals. In this case, the variables are months, days, and hours. In Figure 10, 95% robust confidence intervals have been found, on the one hand, using the α -trimmed mean location estimator and the $s^W(\alpha)$ scale estimator; and, on the other hand, using the biweight estimators $(T_{bi}(c), s_{bi}(c))$.

Table 5. The 95% confidence intervals ($CI_{0.95}$): X_1 (2008), ..., X_{12} (2019).

Variable (Year)	Pairs of Estimators	Lower Limit of the $CI_{0.95}$ ($\mu\text{g}/\text{m}^3$)	Upper Limit of the $CI_{0.95}$ ($\mu\text{g}/\text{m}^3$)	Length of the $CI_{0.95}$ ($\mu\text{g}/\text{m}^3$)
X_1 (2008)	$(T(0.2), s^W(0.2))$	19.8084	20.7331	0.9247
	$(T_{bi}(9), s_{bi}(9))$	16.4104	17.1780	0.7676
X_2 (2009)	$(T(0.2), s^W(0.2))$	22.7226	23.7676	1.0450
	$(T_{bi}(9), s_{bi}(9))$	19.9870	20.8588	0.8718
X_3 (2010)	$(T(0.2), s^W(0.2))$	22.6131	23.6382	1.0251
	$(T_{bi}(9), s_{bi}(9))$	18.5928	19.4419	0.8491
X_4 (2011)	$(T(0.2), s^W(0.2))$	19.4984	20.4428	0.9444
	$(T_{bi}(9), s_{bi}(9))$	17.1457	17.9259	0.7801
X_5 (2012)	$(T(0.2), s^W(0.2))$	21.2993	22.2385	0.9392
	$(T_{bi}(9), s_{bi}(9))$	18.5019	19.2803	0.7784
X_6 (2013)	$(T(0.2), s^W(0.2))$	21.9628	22.9674	1.0046
	$(T_{bi}(9), s_{bi}(9))$	19.5530	20.3653	0.8123
X_7 (2014)	$(T(0.2), s^W(0.2))$	23.2136	24.2042	0.9906
	$(T_{bi}(9), s_{bi}(9))$	19.8670	20.6752	0.8081
X_8 (2015)	$(T(0.2), s^W(0.2))$	24.6840	25.7051	1.0211
	$(T_{bi}(9), s_{bi}(9))$	22.1498	22.9828	0.8329
X_9 (2016)	$(T(0.2), s^W(0.2))$	22.7956	23.8797	1.0840
	$(T_{bi}(9), s_{bi}(9))$	20.6076	21.5038	0.8962
X_{10} (2017)	$(T(0.2), s^W(0.2))$	20.1329	21.0469	0.9140
	$(T_{bi}(9), s_{bi}(9))$	17.4621	18.2149	0.7527
X_{11} (2018)	$(T(0.2), s^W(0.2))$	22.6658	23.7046	1.0388
	$(T_{bi}(9), s_{bi}(9))$	19.3850	20.2467	0.8618
X_{12} (2019)	$(T(0.2), s^W(0.2))$	27.2138	28.3123	1.0985
	$(T_{bi}(9), s_{bi}(9))$	24.3223	25.2218	0.8995

**(a)** $(T(0.2), s^W(0.2))$ 95% confidence intervals: Y_1 = January, ..., Y_{12} = December.**(b)** $(T(0.2), s^W(0.2))$ 95% confidence intervals: Z_1 = Monday, ..., Z_7 = Sunday.**Figure 10.** Cont.

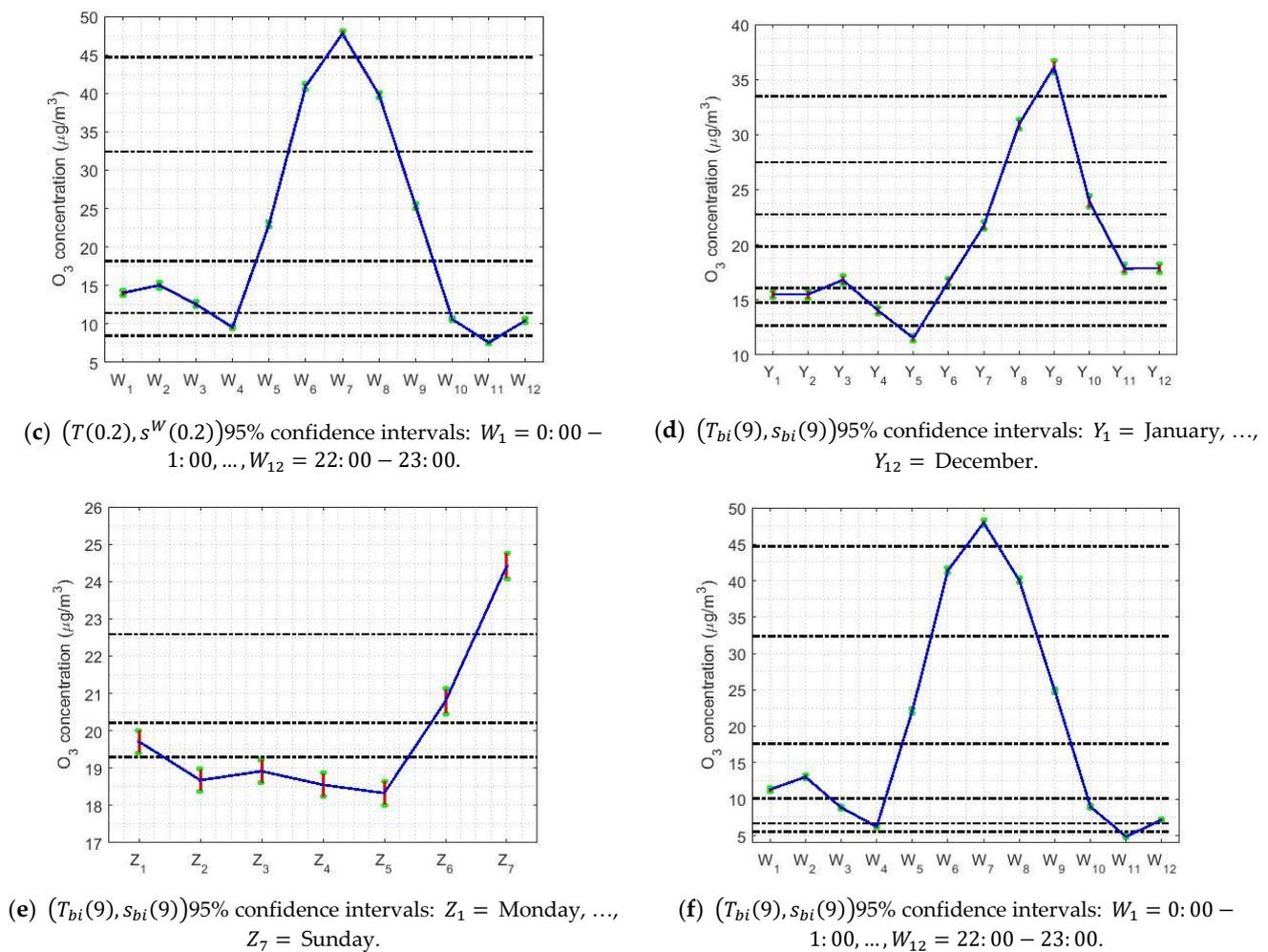


Figure 10. The 95% confidence intervals. The separation between categories is made by the horizontal dashed lines.

In Figure 10a,d, it can be seen that May is when the lowest values (around $12 \mu\text{g}/\text{m}^3$) are reached and the maximum (around $38 \mu\text{g}/\text{m}^3$) is reached in September. This shows that the ozone concentration triples in the summer.

On the other hand, the descent seems to present two steps, one reached in November, when it drops by half (approximately $18 \mu\text{g}/\text{m}^3$), and the other that until March remains stable and then falls until May. Figure 5a shows similar results. In addition, it seems that the analysis of the years yielded wider confidence intervals. Furthermore, it is observed that the median and the width of the intervals are directly related.

Figure 10b,e shows that the analysis for the weeks is in every way analogous to that in Figure 5b, using nonparametric estimators. In short, the maximum is reached on weekends and the values on working days are 33% lower than on weekends. Moreover, there are four categories: (I) Saturday; (II) Sunday; (III) Monday; and (IV) Tuesday, Wednesday, Thursday, and Friday. It is worth mentioning that Mondays and Saturdays are transition categories between working days and Sunday.

Furthermore, Figure 10c,f shows that the analysis for the hours is the same as the one using nonparametric estimators (see Figure 5c). In this case, the maximum is reached at approximately noon, at the moment that most solar lighting occurs, and these confidence intervals have very steep falls that reach two relative minimums. One minimum is around 7:00 and the other is around 21:00, with the deepest fall being seen in this last hour.

It is important to observe in Figure 10c,f that the O_3 concentration goes from approximately $48 \mu\text{g}/\text{m}^3$ to approximately $4 \mu\text{g}/\text{m}^3$, which is a 12-fold reduction. In addition, the

O₃ concentration remains sustained between narrow limits between 21:00 on one day and 6:00 on the next.

Finally, the difference between the pairs of estimators $(T(0.2), s^W(0.2))$ and $(T_{bi}(9), s_{bi}(9))$ is that the separation level of the categories that is achieved by using $(T_{bi}(9), s_{bi}(9))$ is better than that achieved using $(T(0.2), s^W(0.2))$.

5.3. Data Transformation of the Hours at Which the O₃ Concentration Is the Highest to Make Them Fit a Normal Distribution: Analysis by Months and Days of the Week

Sometimes it is important to optimize the data sampling process because there are some applications in which researchers are interested in reduced power consumption when they are using portable devices powered by a battery. This is why, in this part of the paper, some of the categories that have already been established for the months, days, and hours in previous sections are going to be grouped into categories that represent the hours of the highest O₃ concentration.

For the station under study, the analysis carried out previously revealed that the hours of the day at which the ozone concentration levels are highest are between 10:00 and 15:00 (that is, the central hours of the day). Therefore, here we have tried to fit the variables that represent these hours by means of transformations to a normal distribution (that is, $W_6 = 10:00\text{--}11:00$, $W_7 = 12:00\text{--}13:00$, and $W_8 = 14:00\text{--}15:00$). The aforementioned is extremely important because if the desired fittings are achieved, it is possible to use classical statistical inferences to obtain confidence intervals for the mean. Once this is done, the variable changes made can also be undone to obtain confidence intervals for measures of location of the original variables.

In order to do the above, the variables h_i , $i = 1, \dots, 12$, have been introduced to represent the O₃ concentration in the time interval 10:00–15:00 in each of the months. That is, h_1 represents the abovementioned concentration in January, h_2 represents the abovementioned concentration in February, and so on. Table 6 shows the coefficient of skewness of the variables h_i , $i = 1, \dots, 12$, and Figure 11 shows their box plot.

Table 6. Coefficient of skewness of h_i , $i = 1, \dots, 12$.

Variable	h_1	h_2	h_3	h_4	h_5	h_6
Skewness	0.2968	0.4070	0.2679	0.5104	0.4351	0.3874
Variable	h_7	h_8	h_9	h_{10}	h_{11}	h_{12}
Skewness	0.3426	0.4981	0.6277	−0.0688	0.1212	0.2379

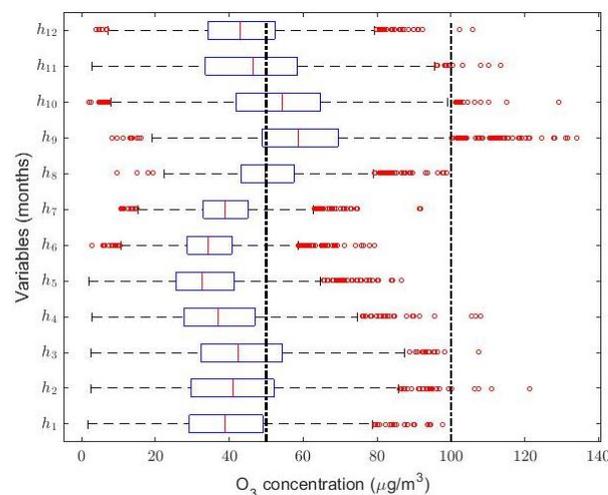


Figure 11. Box plot diagram of the variables h_i , $i = 1, \dots, 12$. The red circles represent the outliers and, according to the IAQ of Quito [23], vertical dashed lines have been used to establish the separation between O₃ concentration levels in $\mu\text{g}/\text{m}^3$: Desirable (i.e., [0, 50)), Acceptable (i.e., [50, 100)), and Caution (i.e., [100, 200)).

Given the shape of the distributions, we used different transformations of variables [68]: linear, logarithmic, and inverse, among others. However, the transformations that have given the best results have been those of the form: $v_i = (a \cdot h_i + b)^q$, $i = 1, \dots, 12$, where a , b and q are positive real numbers. Statistical information on the transformations made to the variables under study is given in Table 7.

Table 7. Statistical information on the transformed variables, v_i , $i = 1, \dots, 12$.

Variable	$v_1 = \sqrt{h_1 + 40}$	$v_2 = h_2$	$v_3 = \sqrt{h_3 + 40}$	$v_4 = \sqrt{h_4 + 20}$	$v_5 = \sqrt{h_5 + 60}$	$v_6 = \sqrt[6]{0.7 \cdot h_6 + 80}$
Mean	8.8780	41.5695	9.0953	7.5570	9.6572	2.0441
Variance	0.7505	291.4980	0.8180	0.9817	0.4336	0.0005
<i>p</i> -Value	0.3336	0.1765	0.8237	0.5638	0.1021	0.0161
Skewness	0.0093	0.4070	−0.0136	0.0799	0.1962	0.0071
Variable	$v_7 = \sqrt[6]{0.5 \cdot h_7 + 65}$	$v_8 = \sqrt{h_8}$	$v_9 = \sqrt[3]{h_9 + 50}$	$v_{10} = (h_{10})^{1.25}$	$v_{11} = h_{11}$	$v_{12} = h_{12}$
Mean	1.9789	7.0760	1.9548	146.9523	46.2482	43.3615
Variance	0.0003	0.6659	0.0018	3615.5855	345.2467	204.7272
<i>p</i> -Value	0.0728	0.4168	0.0431	0.1592	0.6320	0.1607
Skewness	0.0027	0.0599	−0.2071	0.2300	0.1212	0.2379

Therefore, taking into account the information on the *p*-value shown in Table 7, all transformations have been carried out at a 0.01 level of significance. In other words, this means that there is a 1 in 100 chance that we would reject the hypothesis that the transformed variable is Gaussian when it should be accepted. However, despite the fact that transformations to normal distributions have been achieved, there is no pattern that could serve to describe a vast majority of the variables considered. In reality, the process of obtaining the transformation of the variable that can give rise to a normal distribution can be quite laborious, and it is not guaranteed that the desired results will be obtained.

Next, it will be compared whether, once the appropriate transformation has been found, there are significant differences between the classical intervals obtained after the change of variable and those obtained using robust statistics.

The robust confidence intervals and robust centralization estimators, based on the biweight estimators $(T_{bi}(9), s_{bi}(9))$, are represented in Figure 12 to the left of each of the original variables, h_i , $i = 1, \dots, 12$ (on the abscissa axis). Likewise, for the mean of the adjusted normal distribution, the inverse confidence interval is shown on the right side of the original variables, together with the mean of these variables.

Figure 12 shows that both confidence intervals (i.e., the robust confidence interval and the classic confidence interval) are very similar, in terms of both their location and their length. Furthermore, in the case of the variables h_4 and h_9 , the mean is not included within the classical intervals. This shows that the number of outliers makes the distributions of the original variables have heavier tails than the normal distribution. Moreover, as in the other variables (i.e., $h_1, \dots, h_3, h_5, \dots, h_8, h_{10}, \dots, h_{12}$) the mean falls within the classic confidence interval, so the desired estimate can be found by using this type of confidence interval.

In order to conclude the study, an analysis similar to that performed previously will be carried out below, but now said analysis is aimed at studying the hours with the highest O_3 concentration. Therefore, this new analysis will be carried out in the same time slot used previously, depending on the day of the week and distinguishing between working days (i.e., Monday to Friday) and weekends. These three new variables (i.e., *Working days*, *Saturday*, and *Sunday*) have been chosen, because Sunday is the day of the week on which there is the highest O_3 concentration, and Saturday behaves as a transition period between the days people normally have to work and Sunday. The skewness coefficient of the variables *Working days*, *Saturday*, and *Sunday* is shown in Table 8, and their box plot is shown in Figure 13.

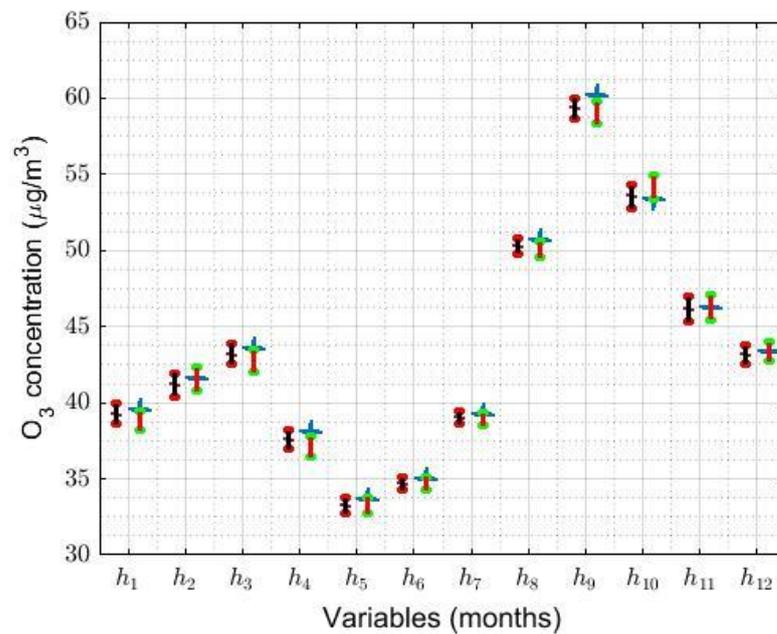


Figure 12. The 95% robust and classic confidence intervals. The robust confidence interval (in black) and robust centralization estimator (at the center of the robust interval) are shown on the left-hand side of $h_i, i = 1, \dots, 12$. In addition, for the mean of the transformed variable (i.e., $v_i, i = 1, \dots, 12$), the inverse transformation of the interval is shown on the right-hand side of h_i (in red). In the i -th classic interval, the mean of the variable h_i is represented by an x (in blue).

Table 8. Coefficient of skewness of the variables *Working days*, *Saturday*, and *Sunday*.

Variable	Working Days	Saturday	Sunday
Skewness	0.4886	0.4596	0.6060

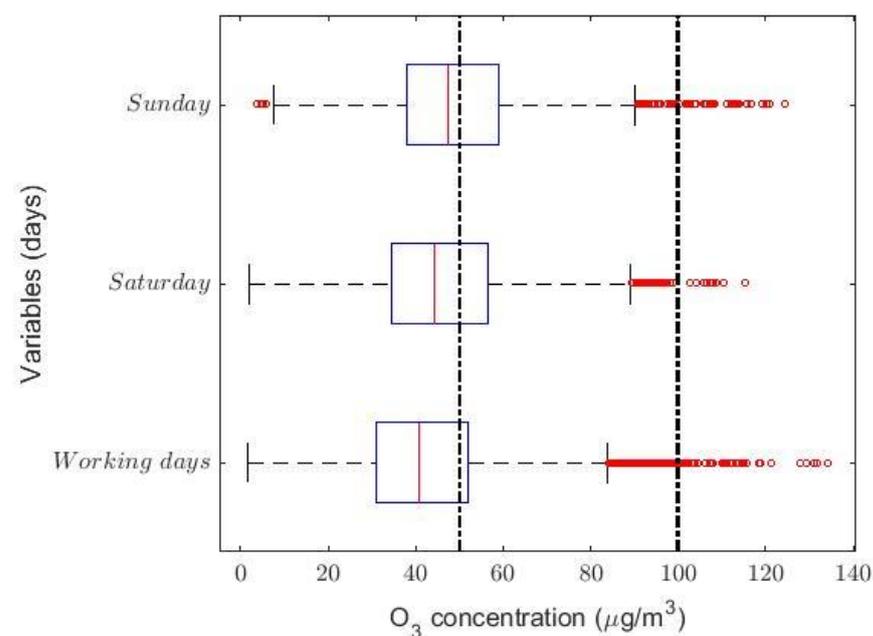


Figure 13. Box plot diagram of the variables *Working days*, *Saturday*, and *Sunday*. The red circles represent the outliers and, according to the IAQ of Quito [23], vertical dashed lines have been used to establish the separation between O_3 concentration levels in $\mu\text{g}/\text{m}^3$: Desirable (i.e., $[0, 50)$), Acceptable (i.e., $[50, 100)$), and Caution (i.e., $[100, 200)$).

In Figure 13, it is observed that the O₃ concentration grows when going from working days to Sunday and that Saturday is a transition category. Moreover, it is observed that these variables seem to have come from heavy-tailed distributions, which can be inferred from the values shown in Table 8 and the box plot diagrams shown in Figure 13. Figure 9 shows statistical information on the transformations made to the variables to make them fit a normal distribution. The transformed variables are shown in Table 9 and all transformations have been carried out at a 0.05 level of significance.

Table 9. Statistical information on the transformed variables, u_i , $i = 1, 2, 3$.

Variable	$u_1 = \sqrt[6]{0.5 \cdot (\text{Working days}) + 30}$	$u_2 = \sqrt{\text{Saturday} + 20}$	$u_3 = \sqrt{\text{Sunday} + 10}$
Mean	1.9225	8.0651	7.6110
Variance	0.0026	1.0587	1.1559
<i>p</i> -Value	0.0524	0.2741	0.3501
Skewness	0.0469	0.0588	0.0976

When grouping the hours and working days, it turns out that the change of variable is quite drastic. Furthermore, Table 9 shows that the *p*-value of the fit of the working days to a normal distribution is the lowest of the three. Additionally, the grouping of the days with a rather artificial, far-fetched transformation, u_1 , and a low *p*-value is an indicator that a combination of different statistical populations is being analyzed.

If a far-fetched transformation and a low *p*-value indicate the existence of a mixture of distributions, then the variables v_6 , v_7 , and v_9 (see Table 7) tell us that the months of June (h_6), July (h_7), and September (h_9) could be analyzed for weeks or fortnights. However, if this is the case, researchers should again start looking for a transformation similar to a normal distribution; then they would have to find the confidence intervals for the mean and, finally, they would have to undo the change of variable. In short, the whole process would have to be repeated again, including what was done up to Section 5.2.

As in Figure 12, Figure 14 shows the robust confidence intervals and robust centralization estimators, based on the biweight estimators ($T_{bi}(9)$, $s_{bi}(9)$), to the left of each of the original variables. On the other hand, for the mean of the adjusted normal distribution, the inverse transformations of the confidence intervals are represented on the right-hand side of the original variables, together with the mean of these variables.

In this case, both types of confidence intervals (i.e., classic and robust ones) are similar in length and, in the case of Saturday and Sunday, they have a nonzero intersection. On the other hand, in the case of working days, the intersection of these intervals is zero, but these intervals do not touch for less than 0.1 µg/m³. Additionally, the mean is not included in any of the classic confidence intervals. Therefore, in this case, the classic confidence intervals are not useful to find the desired estimate.

In view of the results, the following conclusions can be drawn: Obtaining confidence intervals based on robust estimators has turned out to be more efficient than building classic confidence intervals obtained by transforming the variables. As has been shown, in order to obtain classical confidence intervals, transformations of variables must be sought that make them fit a normal distribution. However, these transformations are not homogeneous for all variables and, in some cases, admissible transformations are not obtained either. In fact, in terms of statistics, several of the obtained transformations are rare, artificial, and far-fetched.

Furthermore, something very important that must be highlighted is that, after having done everything that has been done in this subsection (i.e., in Section 5.3), the results obtained in the best of the studied cases are analogous to those previously obtained from the robust estimators of the data provided by the primitive measurements.

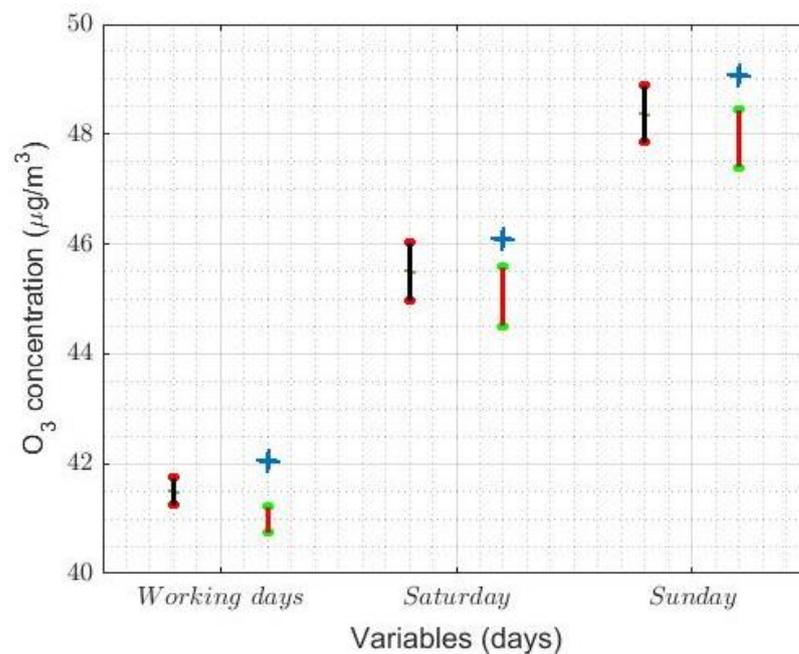


Figure 14. The 95% robust and classic confidence intervals. The robust confidence interval (in black) and robust centralization estimator (at the center of the robust interval) are shown on the left of the variables: *Working days*, *Saturday*, and *Sunday*. Additionally, for the mean of the transformed variable (i.e., u_i , $i = 1, 2, 3$), the inverse transformation of the interval is shown on the right of each untransformed variable (in red). Above the i -th classic interval, the mean of the untransformed variables is represented by an x (in blue).

6. Discussion

A preliminary analysis of the observations yielded that the ozone concentration at Belisario is not a threat to human beings. Most of the samples are at an air pollution level that is considered to be acceptable [23]. Only in specific cases is the level of caution reached, and it happens in a nonsustained manner.

Nevertheless, it should not be forgotten that each city establishes its own criteria for air quality, and what may be acceptable in some cities may not be acceptable in others. However, the truth is that they all share a common goal: to improve the air quality that citizens who live in them breathe every day.

As has happened in previous studies [48–51], the variables chosen for the analysis present a large number of outliers. Therefore, this could indicate that these variables come from heavy-tailed distributions. The foregoing justified the need to combine different types of analyses, in order to explain precisely how the ozone concentration has behaved at the station under study from 2008 to 2019.

The data smoothing process showed a tendency for the O_3 concentration to be entrenched in a range of values. Likewise, for the months, a notable maximum occurred at the end of the summer and, for the hours of the day, the maximum occurred at noon. In this case, we observed notable reductions to the right and left; the concentration of O_3 was then maintained at a stable level until a new rise began. In contrast, the differences between days of the week were seen between weekends and working days, although the ups and downs were not very pronounced.

Once all the previous analyses had been carried out, the authors tried to make the variables under study fit a normal distribution by using variable transformations, because the classic inferential analysis was attempted to be applied. However, the obtained results were not as expected, because when a parametric distribution was attained using a variable transformation for a particular time period, an acceptable fit was not achieved for the other

periods. Therefore, the techniques used to carry out the analysis had to be nonparametric and robust.

Here, it was possible to build different groups of variables by using the WSRT. Moreover, three strata were established to analyze the years by using nonparametric methods. These strata were as follows: one stratum formed by 2019, another by 2015, and the third by the other years. Nevertheless, in the last stratum 2008, 2011, and 2017 can be separated, because these years have the lowest levels of O₃ concentration.

Regarding the analysis carried out for the months using nonparametric methods, seven classes were obtained. However, they were then decreased to four.

Regarding the nonparametric analysis of the days, they were grouped into the last four working days of the week and Monday, Saturday, and Sunday, with Sunday forming one class and Monday and Saturday forming another class. The reason for this classification is the rise in the O₃ concentration on weekends as opposed to working days, and Monday and Saturday being transition categories.

Regarding the hours of the day, many groupings were obtained using nonparametric methods. In this case, a maximum was obtained at noon, the transition categories between the maximum and the minimum corresponded to the hours from 8:00 to 20:00, and in the rest of the hours, from 21:00 on the present day to 7:00 the following day, the O₃ concentration remained fairly unchanged.

Next, robust confidence intervals were built by using location and scale statistics that are robust, because they are highly immune to the influence of extreme observations. The analysis carried out using these confidence intervals showed that the concentration of ozone stayed stable in the range of desirable air pollution level [23], decreasing in 2008, 2011, and 2017, and increasing in 2015 and 2019, with the last year being the one in which the concentration rose much more.

The analysis showed a parallel between location and scale estimates. In short, an increase in the value of the location estimate produced an increase in the value of the scale estimate. One conclusion that can be drawn from this is that the location and scale estimates have been determined by the outliers.

The robust analysis based on confidence intervals allowed us to observe that, in terms of years, a maximum occurred in 2015 and a more pronounced maximum occurred in 2019. In addition, in a more notable way than with the nonparametric analysis, with the robust analysis, 2008, 2011, and 2017 created the lowest value category. Moreover, there were more categories in which the variables were classified when robust intervals are used, because the robust confidence intervals were narrower than the nonparametric ones.

Furthermore, a certain periodicity was observed for the variables corresponding to the months, days, and hours. The analysis of the months allowed us to observe notable increases at the end of the summer, decreases in May and October, and stable behavior in both the first quarter of the year and the last months of it.

The analysis of days allowed us to observe an appreciable difference between the O₃ concentration on working days and the O₃ concentration at weekends. Lastly, in the analysis of the hours of the day, minimum values were observed at 7:00 and 21:00. Additionally, it was observed that the maximum occurred at noon, the transition values were between 7:00 and 21:00, and stable O₃ concentrations were obtained at the rest of the hours.

The existence of maximum and minimum values in the concentration of air pollutants has also been observed in other research [50,51,86]. The reality is that anthropogenic emissions have a great influence on the levels of air pollutants in urban cities.

Other air pollution variables have been shown to have possible periodic behavior [50,51]. Nevertheless, everything seems to indicate that the way in which the data are represented has a great influence on either the appreciation or not of said periodic behavior. The in-depth mathematical modeling of the air pollution variable analyzed in this research is not the objective of this paper. However, this is work to be carried out in future research. The contribution of the present research is that it has been shown that this periodic behavior can be represented by using 95% confidence intervals, both

robust and nonrobust ones. Furthermore, the range of possible values of these signals has been categorized with robust statistical precision, which is in agreement with what was discovered in [50,51].

Through all these analyses, groupings of variables were obtained and differences between categories were found. Here, the quantification of such differences was performed by using confidence intervals.

Finally, it is worth mentioning one more contribution to this paper. The possibility of forming new variables has been analyzed based on the hours of the day that have been shown to have the highest ozone concentrations. Doing this is very important because it allows researchers to optimize the data collection time and therefore save energy when using battery-powered portable measurement equipment. However, the analysis was complex because the new variables formed are different. Therefore, it is difficult to establish predictions since the variables do not come from the same statistical populations. Furthermore, using robust analysis methods was shown to be much more effective than attempting to carry out variable transformations. Here, it was shown that, after having made the variable transformations, the results were analogous at best to those that had already been obtained with the robust methods.

Another thing to keep in mind is that several of the variable transformations achieved are rare, artificial, far-fetched, or unrealistic and therefore do not clarify what the variable does. In addition, the researcher has to first do the robust analysis and then form new variables by grouping together the variables with the highest O₃ concentration values. Therefore, it is not worth doing these additional transformations of variables if the previous robust analysis already allows for obtaining significant results. In other words, it is not advisable to form new variables after having done the robust analysis, if the new results are analogous, in the best of cases, to those already obtained with the robust methods.

7. Conclusions

A robust analysis of ozone concentration was conducted in this paper. This analysis was performed at Belisario station for measurements taken from 1 January 2008 to 31 December 2019. This monitoring station is in Quito, Ecuador, and the set of measurements was transformed into samples taken from several independent variables. The information from these variables was then classified and categorized. In addition, similarities and differences between variables were established with robust statistical precision.

The *raison d'être* of this study was to group the measurements into variables that allowed us to describe the behavior of the O₃ concentration in a more realistic manner. As a matter of fact, people who live in cities are worried about air pollution in terms of hours, weeks, and years, and they need precise information to keep themselves and their relatives safe. The research presented in this paper can be used to know and predict how the O₃ concentration behaves at Belisario station.

In a previous report [23], some general statements about tropospheric ozone levels at Belisario station were made; for the descriptive analysis, the only statistical tools used in [23] were the mean and maximum values. In addition, the report made reference to the main air pollution sources in Quito and presented general environmental issues. However, such a report lacks a robust, rigorous analysis that allows us to study the ozone concentration in a comprehensive manner. Therefore, one possible application of the research presented in this paper is that it can help us to study ozone concentration measurements in a statistical manner and establish robust analysis procedures.

An important fact to highlight is that the city of Quito has greatly progressed in recent years in terms of caring for the environment and improving air quality. However, much remains to be done. The authors of [47] dedicated a few words to this and gave multiple examples of things that still need to be done to improve the air quality in some areas of the city. Moreover, in [50,51] several recommendations were made which, in conjunction with those made in [47], suggested ways to decrease the level of pollution in Quito. The goal is

to inform people about pollution levels in built-up areas and ensure that these pollution levels are low and not harmful to the health of human beings.

The aforementioned entails investment by urban decision-makers, research centers, universities, and so on in measuring equipment that can be located at areas of interest or is portable and battery-powered. Therefore, in order to optimize the sampling process to reduce power consumption, in this research, variables that represented the hours of the day with the highest O₃ concentration were created. These new variables were studied and the results confirmed that, in some cases, it is essential to use robust methods to optimize the data sampling process.

Finally, for some of the variables studied, behaviors that could imply a certain periodicity were observed. However, when the researchers changed the way the data were represented by using different partitions of the time domain, these behaviors were not appreciated. Carrying out the in-depth mathematical analysis of these behaviors was not the objective of this paper, but it is something to consider in future research.

Author Contributions: The methodology was created by W.H. and A.M. In addition, they implemented it and carried out the data analysis and its validation, and wrote the article. V.G.-P., J.L.J.-M. and I.M.C. contributed to the preparation of Section 5.3. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Ecuadorian Corporation for the Development of Research and Academy, Ecuador (under the CEDIA CEPRA XII-2018-13 research Project), the Universidad de Las Americas, Ecuador (under the UDLA ERa.ERI.WHP.18.01 research project), and the Universidad Politecnica de Madrid, Spain.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available in a publicly accessible repository.

Acknowledgments: The authors would like to thank Juan Hernandez, retired chemical engineer and former professor at Instituto Politecnico de Quimica e Informatica Martires de Giron, Havana, Cuba, for helping us to understand tropospheric ozone issues.

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

References

1. U.S. Environmental Protection Agency. Integrated Review Plan for the Review of the Ozone National Ambient Air Quality Standards (EPA-452/R-19-002), Office of Air Quality Planning and Standards, Health and Environmental Impacts Division, Research Triangle Park, NC, and National Center for Environmental Assessment, Research Triangle Park, NC. 2019. Available online: https://www.epa.gov/sites/production/files/2019-08/documents/o3-irp-aug27-2019_final.pdf (accessed on 1 November 2020).
2. Tropospheric Ozone: A Growing Threat, Acid Deposition and Oxidant Research Center, 1182 Sowa, Niigata-City, Niigata, 950-2144, Japan. 2006. Available online: <https://www.acap.asia/wp-content/uploads/Ozone.pdf> (accessed on 1 November 2020).
3. Delgado-Saborit, J.M. Validación e Implementación de Técnicas de Captación Pasiva Para el Estudio de los Niveles y Efectos de Ozono Troposférico y Dióxido de Nitrógeno en un Área Costera Mediterránea. Ph. D. Thesis, Universitat Jaume I, Departament de Química Inorgànica i Orgànica, Autonomous Community of Valencia, Castelló, Spain, 2005.
4. Favre, H.A.; Powell, W.H. *Nomenclature of Organic Chemistry, IUPAC Recommendations and Preferred Names 2013*; The Royal Society of Chemistry, International Union of Pure and Applied Chemistry: Cambridge, UK, 2014.
5. Connelly, N.G.; Damhus, T.; Hartshorn, R.M.; Hutton, A.T. *Nomenclature of Inorganic Chemistry, IUPAC Recommendations 2005*; The Royal Society of Chemistry, International Union of Pure and Applied Chemistry: Cambridge, UK, 2005.
6. CAS REGISTRY—The Gold Standard for Chemical Substance Information. Available online: <https://www.cas.org/support/documentation/chemical-substances> (accessed on 1 November 2020).
7. Baird, C.; Cann, M. *Environmental Chemistry*, 4th ed.; W. H. Freeman and Company: New York, NY, USA, 2008.
8. Matsumi, Y.; Kawasaki, M. Photolysis of Atmospheric Ozone in the Ultraviolet Region. *Chem. Rev.* **2003**, *103*, 4767–4781. [[CrossRef](#)] [[PubMed](#)]
9. Chlorofluorocarbons and Ozone Depletion: A National Historic Chemical Landmark. Available online: <https://www.acs.org/content/acs/en/education/whatischemistry/landmarks/cfcs-ozone.html> (accessed on 1 November 2020).
10. Stevenson, D.S.; Dentener, F.J.; Schultz, M.G.; Ellingsen, K.; van Noije, T.P.C.; Wild, O.; Zeng, G.; Amann, M.; Atherton, C.S.; Bell, N.; et al. Multimodel ensemble simulations of present-day and near-future tropospheric ozone. *J. Geophys. Res. Atmos.* **2006**, *111*, 1–23. [[CrossRef](#)]

11. Mantilla, E.; Castell, N.; Salvador, R.; Azorín, C.; Millán, M.; Miró, J.V.; Juan, L. Ozono Troposférico y Calidad del Aire, Generalitat Valenciana, CEAM. Available online: <http://www.agroambient.gva.es/documents/20549779/161512949/Ozono+troposf%C3%A9rico+y+calidad+del+aire.+Libro+completo/c0ae3c3e-5588-433f-a0b6-daf473a9c0a9> (accessed on 5 November 2020).
12. Hinrichs, R.A.; Kleinbach, M. *Energy: Its Use and the Environment*, 5th ed.; Brooks/Cole, Cengage Learning: Boston, MA, USA, 2013.
13. Finlayson-Pitts, B.J.; Pitts, J.N., Jr. Atmospheric Chemistry of Tropospheric Ozone Formation: Scientific and Regulatory Implications. *Air Waste* **1993**, *43*, 1091–1100. [[CrossRef](#)]
14. Millán, M.M. *Ozone Dynamics in the Mediterranean Basin: A Collection of Scientific Papers Resulting from the MECAPIP, RECAPMA and SECAP Projects*; Air Pollution Research Report 78; CEAM; La Imprenta Comunicación Gráfica, S. L.: Paterna, Valencia, Spain, 2002.
15. World Health Organization. *WHO Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide: Global Update 2005, Summary of Risk Assessment*; World Health Organization: Geneva, Switzerland, 2006.
16. Ozono y Salud. Available online: <https://www.larioja.org/medio-ambiente/es/calidad-aire-cambio-climatico/calidad-aire/ozono-troposferico/ozono-salud> (accessed on 20 September 2020).
17. Ozonators. Available online: <http://encyclopedia.che.engin.umich.edu/Pages/TransportStorage/Ozonators/Ozonators.html> (accessed on 5 November 2020).
18. Hoaglin, D.C.; Mosteller, F.; Tukey, J.W. *Understanding Robust and Exploratory Data Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2000.
19. Maronna, R.A.; Martin, R.D.; Yohai, V.J. *Robust Statistics: Theory and Methods*; John Wiley & Sons: Chichester, West Sussex, UK, 2006.
20. Wilcox, R. *Introduction to Robust Estimation and Hypothesis Testing*, 3rd ed.; Academic Press: Waltham, MA, USA, 2012.
21. Bryson, M.C. Heavy-Tailed Distributions: Properties and Tests. *Technometrics* **1974**, *16*, 61–68. [[CrossRef](#)]
22. Belisario, Quito Air Pollution: Real-time Air Quality Index (AQI). Available online: <https://aqicn.org/city/ecuador/quito/belisario/www.quitoambiente.gob.ec> (accessed on 6 October 2020).
23. Díaz, V. “Informe Calidad del Aire 2017,” Secretaría de Ambiente del Distrito Metropolitano de Quito. Available online: <http://www.quitoambiente.gob.ec/ambiente/index.php/informes#informecalidad-del-aire-017> (accessed on 26 March 2020).
24. Cárdenas, A.M.; Rivera, L.M.; Gómez, B.L.; Valencia, G.M.; Acosta, H.A.; Correa, J.D. Short Communication: Pollution-and-greenhouse gases measurement system. *Measurement* **2018**, *129*, 565–568. [[CrossRef](#)]
25. Astudillo, C.D.; Garza-Castañón, L.E.; Minchala Avila, L.I. Design and Evaluation of a Reliable Low-Cost Atmospheric Pollution Station in Urban Environment. *IEEE Access* **2020**, *8*, 51129–51144. [[CrossRef](#)]
26. Han, Q.; Liu, P.; Zhang, H.; Cai, Z. A Wireless Sensor Network for Monitoring Environmental Quality in the Manufacturing Industry. *IEEE Access* **2019**, *7*, 78108–78119. [[CrossRef](#)]
27. Li, X.; Chen, Q.; Zheng, X.; Li, Y.; Han, M.; Liu, T.; Xiao, J.; Guo, L.; Zeng, W.; Zhang, J.; et al. Effects of ambient ozone concentrations with different averaging times on asthma exacerbations: A meta-analysis. *Sci. Total Environ.* **2019**, *691*, 549–561. [[CrossRef](#)]
28. Song, F.; Sheldon, T.A.; Sutton, A.J.; Abrams, K.R.; Jones, D.R. Methods for Exploring Heterogeneity in Meta-Analysis. *Eval. Health Prof.* **2001**, *24*, 126–151. [[CrossRef](#)]
29. Higgins, J.P.; Thompson, S.G.; Deeks, J.J.; Altman, D.G. Measuring inconsistency in meta-analyses. *BMJ* **2003**, *327*, 557–560. [[CrossRef](#)]
30. Egger, M.; Davey Smith, C.; Schneider, M.; Minder, C. Bias in metaanalysis detected by a simple, graphical test. *BMJ* **1997**, *315*, 629–634. [[CrossRef](#)] [[PubMed](#)]
31. Li, J.; Yin, P.; Wang, L.; Zhang, X.; Liu, J.; Liu, Y.; Zhou, M. Ambient ozone pollution and years of life lost: Association, effect modification, and additional life gain from a nationwide analysis in China. *Environ. Int.* **2020**, *141*, 105771. [[CrossRef](#)] [[PubMed](#)]
32. Tian, Y.; Wu, Y.; Liu, H.; Si, Y.; Wu, Y.; Wang, X.; Wang, M.; Wu, J.; Chen, L.; Wei, C.; et al. The impact of ambient ozone pollution on pneumonia: A nationwide time-series analysis. *Environ. Int.* **2020**, *136*, 105498. [[CrossRef](#)] [[PubMed](#)]
33. Zhao, T.; Tesch, F.; Markevych, L.; Baumbach, C.; Janen, C.; Schmitt, J.; Romanos, M.; Nowak, D.; Heinrich, J. Depression and anxiety with exposure to ozone and particulate matter: An epidemiological claims data analysis. *Int. J. Hyg. Environ. Health* **2020**, *228*, 113562. [[CrossRef](#)] [[PubMed](#)]
34. Zeger, S.L.; Liang, K.-Y. Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics* **1986**, *42*, 121–130. [[CrossRef](#)]
35. Yang, X.; Wu, K.; Wang, H.; Liu, Y.; Gu, S.; Lu, Y.; Zhang, X.; Hu, Y.; Ou, Y.; Wang, S.; et al. Summertime ozone pollution in Sichuan Basin, China: Meteorological conditions, sources and process analysis. *Atmos. Environ.* **2020**, *226*, 117392. [[CrossRef](#)]
36. Baudic, A.; Gros, V.; Sauvage, S.; Locoge, N.; Sanchez, O.; Sarda-Estève, R.; Kalogridis, C.; Petit, J.-E.; Bonnaire, N.; Baisnee, D.; et al. Seasonal variability and source apportionment of volatile organic compounds (VOCs) in the Paris megacity (France). *Atmos. Chem. Phys.* **2016**, *15*, 11961–11989. [[CrossRef](#)]
37. Lyu, X.P.; Chen, N.; Guo, H.; Zhang, W.H.; Wang, N.; Wang, Y.; Liu, M. Ambient volatile organic compounds and their effect on ozone production in Wuhan, central China. *Sci. Total Environ.* **2016**, *541*, 200–209. [[CrossRef](#)]
38. Mousavi, A.; Sowlat, M.H.; Sioutas, C. Diurnal and seasonal trends and source apportionment of redox-active metals in Los Angeles using a novel online metal monitor and Positive Matrix Factorization (PMF). *Atmos. Environ.* **2018**, *174*, 15–24. [[CrossRef](#)]
39. Zeng, P.; Lyu, X.P.; Guo, H.; Cheng, H.R.; Jiang, F.; Pan, W.Z.; Wang, Z.W.; Liang, S.W.; Hu, Y.Q. Causes of ozone pollution in summer in Wuhan, Central China. *Environ. Pollut.* **2018**, *241*, 852–861. [[CrossRef](#)]

40. Xing, J.; Wang, J.; Mathur, R.; Wang, S.; Sarwar, G.; Pleim, J.; Hogrefe, C.; Zhang, Y.; Jiang, J.; Wong, D.C.; et al. Impacts of aerosol direct effects on tropospheric ozone through changes in atmospheric dynamics and photolysis rates. *Atmos. Chem. Phys.* **2017**, *17*, 9869–9883. [[CrossRef](#)] [[PubMed](#)]
41. Zhu, J.; Wang, T.; Bieser, J.; Matthias, V. Source attribution and process analysis for atmospheric mercury in eastern China simulated by CMAQHg. *Atmos. Chem. Phys.* **2015**, *15*, 8767–8779. [[CrossRef](#)]
42. Hollander, M.; Wolfe, D.A.; Chicken, E. *Nonparametric Statistical Methods*, 3rd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2014.
43. Gibbons, J.D.; Chakraborti, S. *Nonparametric Statistical Inference*, 5th ed.; Chapman & Hall/CRC Press, Taylor & Francis Group: Boca Raton, FL, USA, 2011.
44. Al-Momani, T.M.; Al-Nasser, A.D. Statistical Analysis of Air Pollution Caused by Exhaust Gases Emitted from Gasoline Vehicles. *Dirasat Pure Sci.* **2006**, *33*, 93–102.
45. Hernandez, W.; Mendez, A.; Diaz-Marquez, A.M.; Zalakeviciute, R. PM_{2.5} concentration measurement analysis by using nonparametric statistical inference. *IEEE Sens. J.* **2020**, *20*, 1084–1094. [[CrossRef](#)]
46. Hernandez, W.; Mendez, A.; Zalakeviciute, R.; Diaz-Marquez, A.M. Analysis of the information obtained from PM_{2.5} concentration measurements in an urban park. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 6296–6311. [[CrossRef](#)]
47. Hernandez, W.; Mendez, A.; Diaz-Marquez, A.M.; Zalakeviciute, R. Robust Analysis of PM_{2.5} Concentration Measurements in the Ecuadorian Park La Carolina. *Sensors* **2019**, *19*, 4648. [[CrossRef](#)]
48. Hernandez, W.; Mendez, A.; Zalakeviciute, R.; Diaz-Marquez, A.M. Robust Confidence Intervals for PM_{2.5} Concentration Measurements in the Ecuadorian Park La Carolina. *Sensors* **2020**, *20*, 654. [[CrossRef](#)]
49. Hernandez, W.; Mendez, A.; González-Posadas, V.; Jiménez-Martín, J.L. Robust Analysis of the Information Obtained from a Set of 12 Years of SO₂ Concentration Measurements. *IEEE Access* **2020**, *8*, 144976–144992. [[CrossRef](#)]
50. Hernandez, W.; Mendez, A.M. Robust Estimation of Carbon Monoxide Measurements. *Sensors* **2020**, *20*, 4958. [[CrossRef](#)]
51. Hernandez, W.; Mendez, A.M. Twelve-Year Analysis of NO₂ Concentration Measurements at Belisario Station (Quito, Ecuador) Using Statistical Inference Techniques. *Sensors* **2020**, *20*, 5831. [[CrossRef](#)]
52. Wang, S.; Van der A, R.J.; Stammes, P.; Wang, W.; Zhang, P.; Lu, N.; Fang, L. Carbon Dioxide Retrieval from TanSat Observations and Validation with TCCON Measurements. *Remote Sens.* **2020**, *12*, 2204. [[CrossRef](#)]
53. Baire, M.; Melis, A.; Lodi, M.B.; Tuveri, P.; Dachena, C.; Simone, M.; Fanti, A.; Fumera, G.; Pisanu, T.; Mazzarella, G. A wireless sensors network for monitoring the carasau bread manufacturing process. *Electronics* **2019**, *8*, 1541. [[CrossRef](#)]
54. Tang, C.-S.; Wu, T.-Y.; Chuang, K.-J.; Chang, T.-Y.; Chuang, H.-C.; Candice Lung, S.-C.; Chang, L.-T. Impacts of in-cabin exposure to size-fractionated particulate matters and carbon monoxide on changes in heart rate variability for healthy public transit commuters. *Atmosphere* **2019**, *10*, 409. [[CrossRef](#)]
55. Cavaliere, A.; Carotenuto, F.; Di Gennaro, F.; Gioli, B.; Gualtieri, G.; Martelli, F.; Matese, A.; Toscano, P.; Vagnoli, C.; Zaldei, A. Development of Low-Cost Air Quality Stations for Next Generation Monitoring Networks: Calibration and Validation of PM_{2.5} and PM₁₀ Sensors. *Sensors* **2018**, *18*, 2843. [[CrossRef](#)] [[PubMed](#)]
56. Munir, S. Analysing Temporal Trends in the Ratios of PM_{2.5}/PM₁₀ in the UK. *Aerosol Air Qual. Res.* **2017**, *17*, 34–48. [[CrossRef](#)]
57. Pandya, S.; Ghayvat, H.; Sur, A.; Awais, M.; Kotecha, K.; Saxena, S.; Jassal, N.; Pingale, G. Pollution Weather Prediction System: Smart Outdoor Pollution Monitoring and Prediction for Healthy Breathing and Living. *Sensors* **2020**, *20*, 5448. [[CrossRef](#)]
58. Feenstra, B.; Papapostolou, V.; Der Boghossian, B.; Cocker, D.; Polidori, A. Development of a Network of Accurate Ozone Sensing Nodes for Parallel Monitoring in a Site Relocation Study. *Sensors* **2020**, *20*, 16. [[CrossRef](#)]
59. Ferrer-Cid, P.; Barcelo-Ordinas, J.M.; Garcia-Vidal, J.; Ripoll, A.; Viana, M. A comparative Study of Calibration Methods for Low-Cost Ozone Sensors in IoT Platforms. *IEEE Internet Things J.* **2019**, *6*, 9563–9571. [[CrossRef](#)]
60. Alhasa, K.M.; Mohd Nadzir, M.S.; Olalekan, P.; Latif, M.T.; Yusup, Y.; Iqbal Faruque, M.R.; Ahamad, F.; Abd. Hamid, H.H.; Aiyub, K.; Md Ali, S.H.; et al. Calibration Model of a Low-Cost Air Quality Sensor Using an Adaptive Neuro-Fuzzy Inference System. *Sensors* **2018**, *18*, 4380. [[CrossRef](#)]
61. Gaikovich, K.P.; Kropotkina, E.P.; Rozanov, S.B. Statistical Analysis of 1996–2017 Ozone Profile Data Obtained by Ground-Based Microwave Radiometry. *Remote Sens.* **2020**, *12*, 3374. [[CrossRef](#)]
62. What's Best for Measuring Ozone, Electrochemical Sensors of HMOS? Available online: <https://www.aeroqual.com/electrochemical-sensors-hmos> (accessed on 30 November 2020).
63. Model 49C: UV Photometric O₃ Analyzer, Instruction Manual, P/N 9999. Available online: <https://old.faam.ac.uk/index.php/faam-documents/science-instruments/56-te49c-user-manual/file> (accessed on 20 September 2020).
64. Model 49i: UV Photometric O₃ Analyzer, Instruction Manual, Part number 102434-00. Available online: <https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FEPM%2Fmanuals%2FEPM-manual-Model%252049i.pdf&title=TW9kZWwgNDlpIC0gVXNlciBNYW51YWw> (accessed on 20 September 2020).
65. *List of Designated Reference and Equivalent Methods*; United States Environmental Protection Agency, Center for Environmental Measurements & Modeling, Air Methods & Characterization Division (MD-D205-03): Research Triangle Park, NC, USA, 2020.
66. Ghazaly, C.; Hébrant, M.; Langlois, E.; Castel, B.; Guillemot, M.; Etienne, M. Real-Time Ozone Sensor Based on Selective Oxidation of Methylene Blue in Mesoporous Silica Films. *Sensors* **2019**, *19*, 3508. [[CrossRef](#)] [[PubMed](#)]
67. Khan, S.; Newport, D.; Le Calvé, S. Gas Detection Using Portable Deep-UV Absorption Spectrophotometry: A Review. *Sensors* **2019**, *19*, 5210. [[CrossRef](#)] [[PubMed](#)]

68. Pang, X.; Shaw, M.D.; Lewis, A.C.; Carpenter, L.J.; Batchellier, T. Electrochemical ozone sensors: A miniaturised alternative for ozone measurements in laboratory experiments and air-quality monitoring. *Sens. Actuators B* **2017**, *240*, 829–837. [[CrossRef](#)]
69. Petani, L.; Koker, L.; Herrmann, J.; Hagenmeyer, V.; Gengenbach, U.; Pylatiuk, C. Recent Developments in Ozone Sensor Technology for Medical Applications. *Micromachines* **2020**, *11*, 624. [[CrossRef](#)]
70. Gavazza, A.; Marchegiani, A.; Rossi, G.; Franzini, M.; Spaterna, A.; Mangiaterra, S.; Cerquetella, M. Ozone Therapy as a Possible Option in COVID-19 Management. *Front. Public Health* **2020**, *8*, 417. [[CrossRef](#)]
71. Cattel, F.; Giordano, S.; Bertiond, C.; Lupia, T.; Corcione, S.; Scaldaferrri, M.; Angelone, L.; De Rosa, F.G. Ozone therapy in COVID-19: A narrative review. *Virus Res.* **2021**, *291*, 198207. [[CrossRef](#)]
72. David, M.; Haniff Ibrahim, M.; Mahdaliza Idrus, S.; Izam Azmi, A.; Hafizah Ngajikin, N.; Ching En Marcus, T.; Yaacob, M.; Rashidi Salim, M.; Abdul Aziz, A. Progress in Ozone Sensors Performance: A Review. *J. Teknol.* **2015**, *73*, 23–29. [[CrossRef](#)]
73. Lal Kishore, K. *Electronic Measurements & Instrumentation*; Dorling Kindersley (India) Pvt. Ltd., Licensees of Pearson Education in South Asia: New Delhi, India, 2010.
74. Oppenheim, A.V.; Willsky, A.S.; Nawa, S.H. *Signals & Systems*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 1997.
75. Zhou, K.; Doyle, J.C.; Glover, K. *Robust and Optimal Control*; Prentice Hall: Englewood Cliffs, NJ, USA, 1996.
76. *Quality Assurance Handbook for Air Pollution Measurement Systems, Volume II. Ambient Air Quality Monitoring Program*; EPA-454/B-17-001; U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Air Quality Assessment Division: Research Triangle Park, NC, USA, 2017.
77. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*, 5th ed.; Jhon Wiley & Sons: Hoboken, NJ, USA, 2016.
78. Brockwell, P.J.; Davis, R.A. *Introduction to Time Series and Forecasting*, 2nd ed.; Springer: New York, NY, USA, 2002.
79. Papoulis, A.; Unnikrishna Pillai, S. *Probability, Random Variables, and Stochastic Processes*, 4th ed.; McGraw-Hill Higher Education: New York, NY, USA, 2002.
80. Hampel, F.R. The Influence Curve and its Role in Robust Estimation. *J. Am. Stat. Assoc.* **1974**, *69*, 383–393. [[CrossRef](#)]
81. Tukey, J.W. *Exploratory Data Analysis*; Addison-Wesley: Reading, Boston, MA, USA, 1977.
82. Croux, C.; Rouseeuw, P.J. A Class of High-Breakdown Scale Estimators Based on Subranges. *Commun. Stat. Theory Methods* **1992**, *21*, 1935–1951. [[CrossRef](#)]
83. Rock, N.M.S. ROBUST: An Interactive FORTRAN-77 Package for Exploratory Data Analysis Using Parametric, Robust and Nonparametric Location and Scale Estimates, Data Transformations, Normality Tests, and Outlier Assessment. *Comput. Geosci.* **1987**, *13*, 463–494. [[CrossRef](#)]
84. Dixon, W.J.; Tukey, J.W. Approximate Behavior of the Distribution of Winsorized t (Trimming/Winsorization 2). *Technometrics* **1968**, *10*, 83–98. [[CrossRef](#)]
85. Mosteller, F.; Tukey, J.W. *Data Analysis and Regression: A Second Course in Statistics*; Addison-Wesley: Reading, MA, USA, 1977.
86. Lun Fung, P.; Zaidan, M.A.; Sillanpää, S.; Kousa, A.; Niemi, J.V.; Timonen, H.; Kuula, J.; Saukko, E.; Luoma, K.; Petäjä, T.; et al. Input-Adaptive Proxy for Black Carbon as a Virtual Sensor. *Sensors* **2020**, *20*, 182. [[CrossRef](#)] [[PubMed](#)]