atmosphere

MDPI

Article

A Spatio-Temporal Visualization Approach of PM_{10} Concentration Data in Metropolitan Lima

Alexandra Abigail Encalada-Malca [1], Javier David Cochachi-Bustamante [1], Paulo Canas Rodrigues [2], Rodrigo Salas [3] and Javier Linkolk López-Gonzales [1,4,*]

[1] Facultad de Ingeniería y Arquitectura, Universidad Peruana Unión, Lima 15, Peru; alexandraencalada@upeu.edu.pe (A.A.E.-M.); javiercochachi@upeu.edu.pe (J.D.C.-B.)
[2] Departament of Statistics, Federal University of Bahia, Salvador 40170-110, Brazil; paulocanas@gmail.com
[3] Escuela de Ingeniería C. Biomédica, Universidad de Valparaíso, Valparaíso 2362905, Chile; rodrigo.salas@uv.cl
[4] Instituto de Estadística, Universidad de Valparaíso, Valparaíso 2360102, Chile
* Correspondence: javierlinkolk@gmail.com

Abstract: Lima is considered one of the cities with the highest air pollution in Latin America. Institutions such as DIGESA, PROTRANSPORTE and SENAMHI are in charge of permanently monitoring air quality; therefore, the air quality visualization system must manage large amounts of data of different concentrations. In this study, a spatio-temporal visualization approach was developed for the exploration of data of the PM_{10} concentration in Metropolitan Lima, where the spatial behavior, at different time scales, of hourly concentrations of PM_{10} are analyzed using basic and specialized charts. The results show that the stations located to the east side of the metropolitan area had the highest concentrations, in contrast to the stations located in the center and north that reported better air quality. According to the temporal variation, the station with the highest average of biannual and annual PM_{10} was the HCH station. The highest PM_{10} concentrations were registered in 2018, during the summer, highlighting the month of March with daily averages that reached 435 $\mu g/m^3$. During the study period, the CRB was the station that recorded the lowest concentrations and the only one that met the Environmental Quality Standard for air quality. The proposed approach exposes a sequence of steps for the elaboration of charts with increasingly specific time periods according to their relevance, and a statistical analysis, such as the dynamic temporal correlation, that allows to obtain a detailed visualization of the spatio-temporal variations of PM_{10} concentrations. Furthermore, it was concluded that the meteorological variables do not indicate a causal relationship with respect to PM_{10} levels, but rather that the concentrations of particulate material are related to the urban characteristics of each district.

Keywords: air quality; PM_{10}; spatio-temporal visualization; particulate matter; dynamic time warping

Citation: Encalada-Malca, A.A.; Cochachi-Bustamante, J.D.; Canas, R.P.; Salas, R.; López-Gonzales, J.L. A Spatio-Temporal Visualization Approach of PM_{10} Concentration Data in Metropolitan Lima. *Atmosphere* **2021**, *12*, 609. https://doi.org/10.3390/atmos12050609

Academic Editor: Pasquale Avino

Received: 4 April 2021
Accepted: 27 April 2021
Published: 7 May 2021

1. Introduction

Emissions released into the atmosphere alter its composition and affect air quality, causing significant repercussions in areas such as the environment and health. Air pollution is linked to the presence of substances that reach concentrations higher than their normal environments [1], causing an ecological imbalance [2]. These episodes produce potentially harmful problems for human health, animals and vegetation. In urban places, most of the contamination is generated by emissions from automobiles and industries [1,3]. These emissions have an environmental impact due to the transport of pollutants, even to great distances from the emission points, due to the action of the wind and other factors such as turbulence [4,5], which is why affected sectors are not exclusively industrial [6]. According to epidemiological studies, one of the main risk factors for the global burden of mortality and morbidity is air pollution [7]. It is estimated that in the coming decades it could become the largest environmental cause of premature death [8], for which global

mitigation policies have been implemented by deploying devices that control air quality on a daily basis in their territories [9].

In Latin America (LA) there is a high rate of urbanization, comprising 72% of the population [8]. The main causes of air pollution in the cities of Latin America and the Caribbean (LAC) are land transportation, electricity generation and industrial production [10,11].

During the 2010-2014 period, most of the official air quality monitoring stations in seventeen LAC countries recorded mean annual PM_{10} values significantly higher than the limit of 20 $\mu g/m^3$ established in guidelines of the World Health Organization (WHO) [12]. According to the 2019 World Air Quality Report, nine of the twelve considered LAC countries reported an annual mean of $PM_{2.5}$ higher than the threshold established by the WHO, with Chile and Peru being the countries with the highest average annual exposure to pollution [13].

Peru is a highly urban country, reporting a higher density in Lima, considered one of the most polluted cities in LA [8]. The Department (second-level administrative subdivisions of Peru) of Lima is located on the central coast; it has a humid climate and a geography characterized by the Andes that reach the seashore at high altitudes. Its main source of air pollution is transportation units, as well as stationary sources such as industries, shops, and restaurants [14–16]. In some areas the levels of air pollution are very high due to the traffic congestion of public transport, which mostly use diesel fuel, are old and do not have adequate maintenance [17,18]. Likewise, it is one of the cities with the highest frequency of asthma in children, caused by particulate matter from emissions from the vehicle fleet [19].

The minimum values of PM_{10} during the period 2001–2014 were recorded on Sundays, coinciding with the decrease in anthropogenic activities and the flow of vehicular transport that occurs on those days [20]. In a report by the Japan International Cooperation Agency (JICA), it was stated that, in 2012, the number of trips made per day in motorized vehicles in Lima was 16.9 million [21]. Over the years, this number has increased, increasing also the vehicle gas emissions [8]. Several strategies were implemented, such as the elimination of lead, gasoline and the reduction of the sulfur content in diesel, that serve to mitigate this problem [22]. The permanent monitoring of air quality is done by institutions such as DIGESA, PROTRANSPORTE and SENAMHI [23].

In this sense, monitoring combined with other objective evaluation techniques helps to fully define and understand the exposure of the population to the environment [6]. The visual visualization of the data takes a dynamic approach compared to raw data tables, and provides detailed information on data structure [24], ensuring that the results are directly transmitted to the general public [9]. This technique can transform data into intuitive graphical images, showing efficient visual analysis of air quality data, thereby determining the location of sources of pollution and assessing which areas are more affected, in order to take necessary actions for improvement [3]. The air quality display system must handle large amounts of spatial data of the different concentrations. In addition, they must be viewed in an agile way to detect trends, correlations and alterations in air pollution [25,26], becoming a solution that allows complex data to be simplified and better understood for timely decision-making [27].

In Metropolitan Lima, studies such as those by Valdivia [20] and Silva et al. [28] have applied air quality data visualization techniques in periods between 2001 to 2014 and 2010 to 2015, respectively. Likewise, studies such as those of Reátegui-Romero et al. [29] and Sánchez-Ccoyllo et al. [30] have applied particulate matter estimation techniques in Metropolitan Lima. These modeling techniques allow complementing the application of air quality data visualization.

This article proposes a spatio-temporal visualization approach to explore PM_{10} concentration data in the Metropolitan Lima area, through graphs with increasingly specific time periods (year, month, day, hour) according to their relevance and statistical analysis for better interpretation. This will allow to obtain a complete, realistic and detailed visualization of the concentrations and spatio-temporal variations of PM_{10}.

2. Methods

The proposed methodology is based on the study developed by Li et al. [27] for the exploration of air pollution data, and has been adapted to the PM_{10} concentration data in Metropolitan Lima. The workflow includes five steps that start with access to the database, then its preliminary analysis in step 2, from which the hypotheses of step 3 are formulated, which are verified in step 4, with the application and visualization conducted in step 5. In addition, the proposed approach for each of the steps and their work schemes are described in the following subsections.

2.1. Database

The first step of the workflow consists of data access and storage. The data comprises a period of 2 years (from 1 January 2017 to 31 December 2018); and includes time series (year, month, day, hour) of PM_{10} concentration ($\mu g/m^3$), temperature (ºC), humidity (%), wind speed (m/s) and wind direction; coming from 5 air quality monitoring stations of the "Servicio Nacional de Meteorología e Hidrología" (SENAMHI): Ate (ATE), Jesús María (CDM), Carabayllo (CRB), Huachipa (HCH) and San Martín de Porres (SMP). The geographical distribution of the stations is presented in Figure 1, being two stations located in North Lima, two stations located in East Lima and one station located in Center Lima.

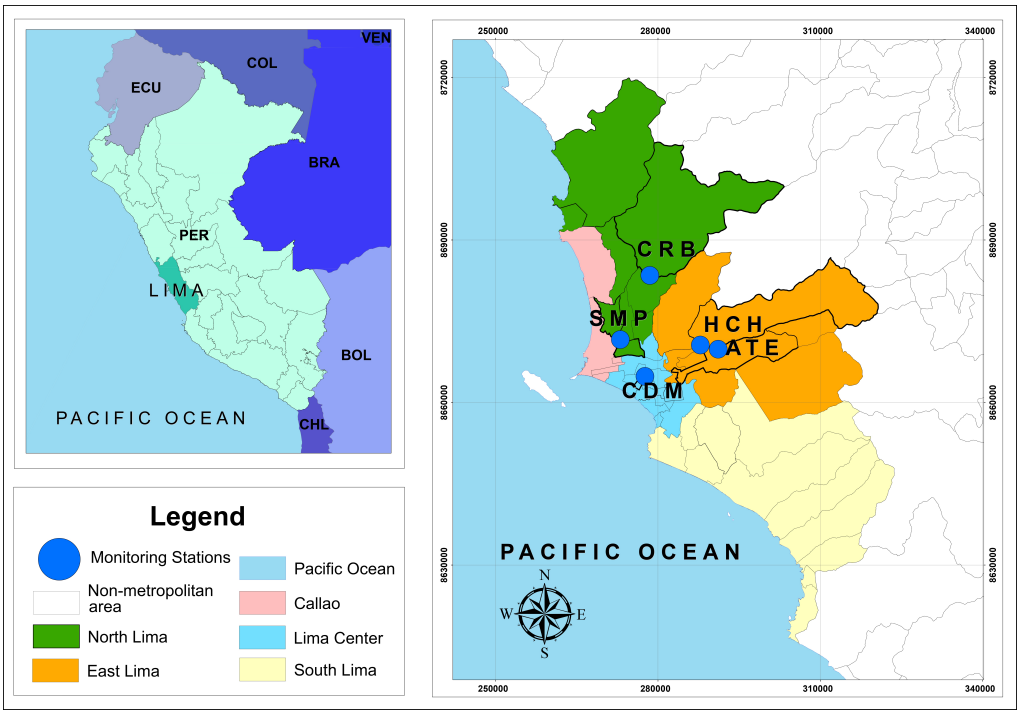

Figure 1. Geo-spatial locations of the five monitoring stations: Ate (ATE), Jesús María (CDM), Carabayllo (CRB), Huachipa (HCH) and San Martín de Porres (SMP).

2.2. Preliminary Analysis

The preliminary analysis aimed at performing the cleaning of the data and imputation of missing values, as well as a general report with basic plots such as heat tables, line and bar charts. The main findings of this stage give an important contribution to better understand patterns in the data and to contribute to decision-making.

2.2.1. Data Imputation

The imputation of the missing data was done with the R package MICE, where multiple imputation was considered by using the Fully Conditionally Specification [31]. The imputation was done for each variable separately and, for the variables under study (PM_{10}, temperature, relative humidity and wind speed), predictive mean matching was

used, a semi-parametric and versatile method that focuses on continuous data, which allows the imputed values to coincide with any of the observed values in each variable, that is, it will not produce imputations outside the range of observed data. The percentages of missing data was between 12% and 27%. Other stations with much higher percentages of missing values were not considered.

2.2.2. Preliminary Visualization

Line and bar charts together with heat tables were used for preliminary visualizations to better understand the behavior of the data.

Figure 2 presents the sequence of steps developed for the preliminary visualization of the data. The first step is the data exploration, which consists in the identification of the existing variables of the original data. In the second step, the variables were analysed to decide the types of plots to be generated. In that sense, variables such as: latitude, longitude, day of the week, hour, season of the year and concentration level were added and organized. For this study it was considered that the seasons of the year in Lima are: summer (22 December to 21 March), fall (22 March to 21 June), winter (22 June to 22 September) and spring (23 September to 21 December). In the third step, the charts are designed and obtained from the most general to the most specific. Six types of graphs are presented for all-station PM_{10} concentration data, eleven types of graphs for single-station PM_{10} concentration data, and four types of graphs for data for each meteorological variable per station. For the graphs of meteorological variables, a smaller quantity has been considered, taking into account that their main variations are seasonal and hourly; these graphs also include the behavior of the recorded PM_{10} concentrations, in order to obtain a visualization with respect to the meteorology. This approach proposes several types of graphs and with increasingly specific time periods, to better observe the most critical values at different time-scales. Likewise, this type of approach allows to focus on the highest recorded values and analyze their possible causes and/or consequences. Steps 4 and 5 of Figure 2 correspond to the selection of graphs, their analysis and interpretation. The selected plots depend on the user and the objective of the analysis.

2.3. Hypothesis

The preliminary analysis describes the important findings obtained from the basic plots. Some of these findings need to be verified, by considering specific hypotheses. Three hypotheses were raised: The first, regarding PM_{10} concentrations, proposes that the stations present an optimal coincidence between the time series. The second affirms that the meteorological variables have a causal/correlational relationship with respect to PM_{10} concentrations. Finally, the third hypothesis, based on the detailed analysis of the station with the highest concentration of PM_{10}: HCH, suggests that there is a positive correlation between temperature and PM_{10} concentration at the HCH station.

2.4. Verification

The hypotheses are verified through visual analytics, using scatter charts, boxplots and graphs from statistical analysis using the Dynamic Time Warping (DTW) algorithm. The latter allows to optimally align two time series [32]. In addition, clustering analysis is useful for detecting similar patterns in the data that comprises multiple time series [33,34]. The DTW uses a dynamic programming approach that compresses the two time series to minimize the difference between them as much as possible, thus ensuring that the two time series are ordered through a deformation path [35].

Geo-located wind roses, considering the direction and speed of the wind, were also generated to complement the results. Finally, geo-visualization charts were used to represent the spatial distribution of the air pollution data.

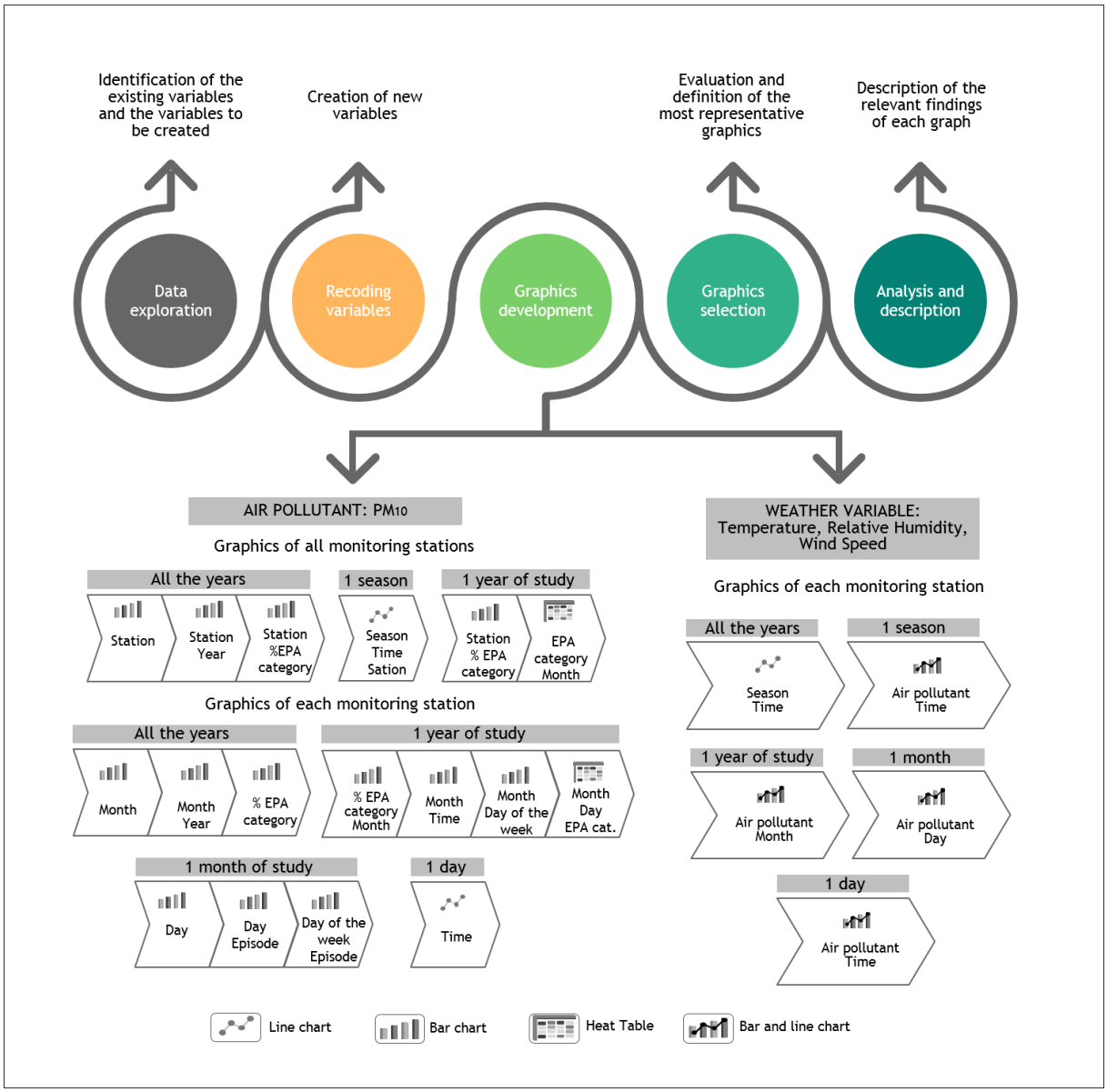

Figure 2. Proposal for a preliminary analysis of the data visualization from the air quality monitoring stations in Metropolitan Lima. The proposal is developed in five steps that range from the exploration of the data to the analysis and description of the plots.

2.5. Application

The applications of the study are based in a descriptive analysis and visualization. In this last block, the relevant findings of steps 2 and 4 are considered, as well as the subsequent analysis that will complement the study according to the approach. It is important to consider that, for the application, it can be linked with other systems, providing a basis for a comprehensive model of a phenomenon. The proposed methodology has been applied in the spatio-temporal visualization of the air quality data of Metropolitan Lima. Likewise, this approach could also be extrapolated to other large cities.

3. Results and Discussion

Initially, descriptive analysis of the data were carried out, covering the period between 1 January 2017 and 31 December 2018, with five monitoring stations for PM_{10}. Table 1 shows the descriptive statistics for each station, showing an asymmetry in the data. In this sense, both the histograms and the boxplots, showed in Figures 3 and 4, represent the best option for visualizing PM_{10} asymmetry, allowing to evaluate the behavior in each monitoring

station. Meanwhile, the PM_{10} pollutant has a heavy-tailed probability distribution (positive asymmetry), which is evidenced by the appearance of high-pollution observations.

Table 1. Summary statistics - Stations of Monitoring PM_{10}.

Stations	Min.	Max.	1st Qu.	3rd Qu.	Median	Mean ± DS	Var.	Skew.	Kurt.
CRB	5.44	488.02	31.49	58.45	198.31	48.69 ± 28.39	806.03	3.24	22.27
SMP	7.77	426.80	61.95	105.10	142.50	86.05 ± 35.73	1276.41	1.00	2.86
CDM	6.08	463.60	35.84	63.45	145.50	52.30 ± 24.61	605.54	2.30	18.25
ATE	6.41	931.00	82.90	148.00	421.90	121.56 ± 60.30	3635.75	2.08	11.07
HCH	5.21	974.00	62.10	176.50	138.40	130.03 ± 91.68	8404.34	1.53	4.89

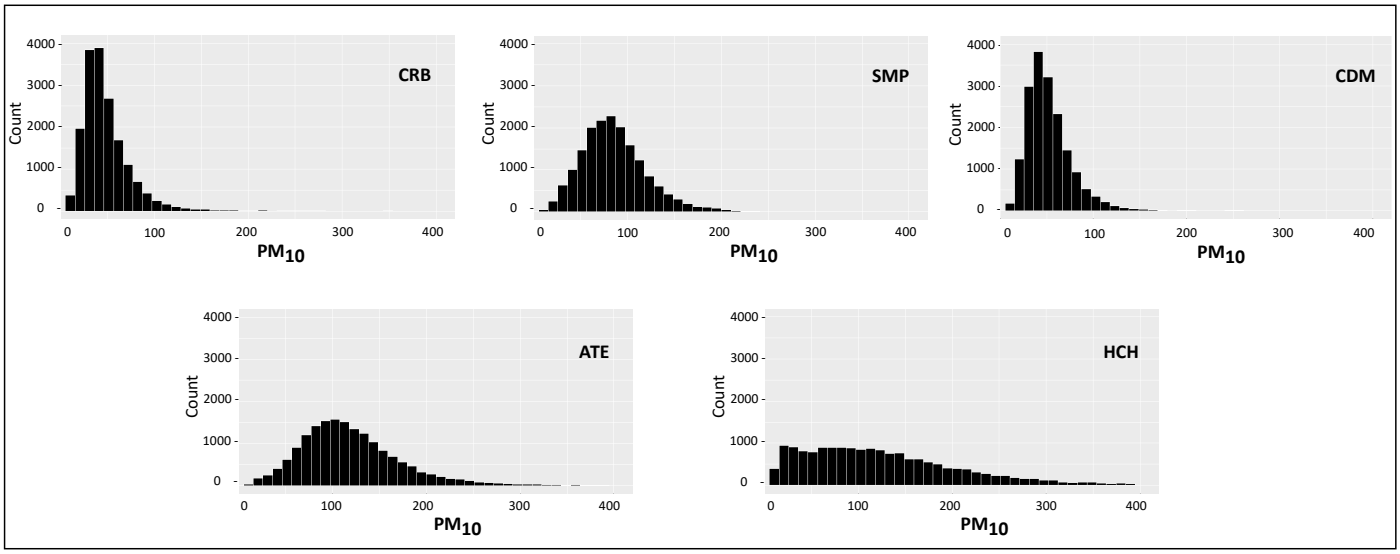

Figure 3. Histogram of PM_{10} for each station. The stations with the highest skewness are CRB, CDM and ATE, showing an heavy tail towards the right with respect to the mean of PM_{10}.

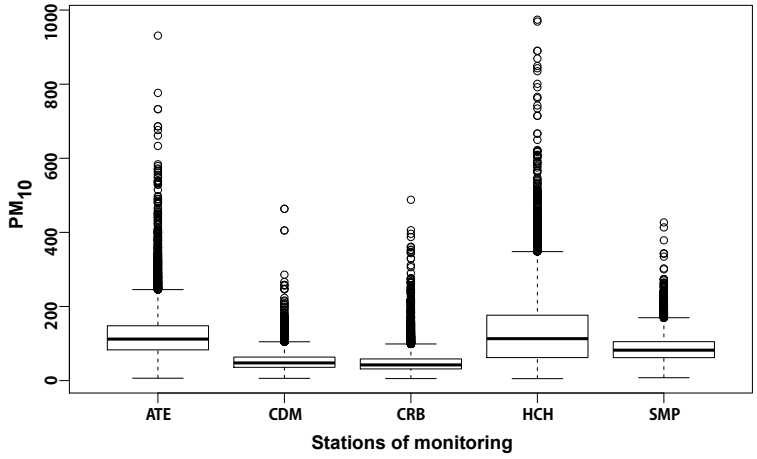

Figure 4. Boxplots for PM_{10} in each station.

3.1. Temporal Variation of PM_{10}

Figures 5–7 show the temporal visualisation of PM_{10} concentration. For a more specific analysis, the cases in which the concentration levels were higher, have been considered. Figure 5, shows the behavior of the data in the five stations. Figure 5a shows the overall averages of the PM_{10} concentrations (from 1 January 2017 to 31 December 2018). In this period, all the stations exceed the limits proposed by the WHO for the maximum concentrations of

PM_{10} in a year. In addition, all stations, except for CRB, exceed the annual arithmetic mean of PM_{10} established in the ECA. Figure 5b compares the annual averages of PM_{10} for 2017 and 2018 in each station. In this figure it is visible that the difference between the values from one year to another does not exceed 10 $\mu g/m^3$ in all stations. However, the annual average of PM_{10} in 2017 from the CRB station exceeds the concentration established in the ECA for annual average compared to that is shown in Figure 5a. In Figure 5c, the categories of Air Quality Index (AQI) from Table 2 were considered, being the percentage of hourly concentrations during the study period shown in the bar chart for each of the categories by station.

Table 2. Breakpoints for the Air Quality Index—AQI [36].

Category	PM_{10} ($\mu g/m^3$) 24-h	AQI	Color
Good	0–54	0–50	Green
Moderate	55–154	51–100	Yellow
Unhealthy for sensitive groups	155–254	101–150	Orange
Unhealthy	255–354	151–200	Red
Very Unhealthy	355–424	201–300	Purple
Hazardous	425–604	301–400	Sangria

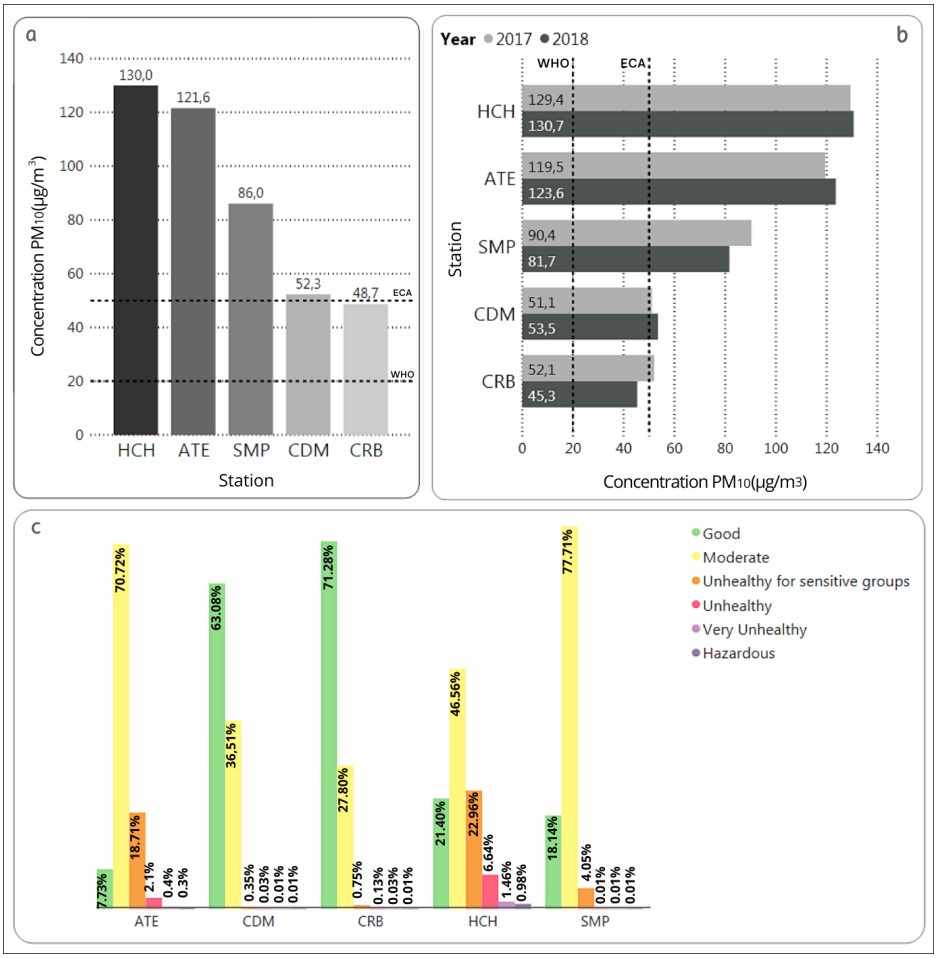

Figure 5. (**a**) Average of the PM_{10} concentrations in the period that extends from 1 January 2017 to 31 December 2018. The horizontal lines represent the limit values for annual averages of PM_{10} proposed in the ECA in Peru and by the WHO. (**b**) Distribution of the annual averages concentration of PM_{10} by station. The vertical lines represent the limit values for annual averages of PM_{10} proposed in the ECA and by the WHO. (**c**) Distribution of PM_{10} concentration levels by station.

Figure 6. (**a**) Monthly comparison of PM_{10} concentration in the period 2017–2018 for the HCH station. (**b**) Daily variation of PM_{10} concentration per month in the year 2018 for the HCH station. (**c**) 2018 heat table at the HCH station. (**d**) Hourly variation of the PM_{10} concentration per month in 2018 for the HCH station.

The HCH station records the highest percentages of data in the categories: Unhealthy for sensitive people (22.96%), unhealthy (6.64%), very unhealthy (1.46%) and hazardous (0.98%); these percentages are equivalent to 4022, 1164, 255 and 171 observations/hours of PM_{10} concentration, respectively, of a total of 17,520 observations.

According to the analysis of the multiannual hourly variation of PM_{10} for the monitoring stations in summer, fall, winter and spring, the highest concentrations of PM_{10} were recorded during the summer. In the summer, HCH and ATE are the stations that recorded the highest PM_{10} concentration values; in the case of HCH, 100% of its values exceed the ECA in a range of 124–203 µg/m^3. In the period 2010–2015 it was also during the summer that the highest concentrations were recorded in Metropolitan Lima [28]. During the period 2007–2014, the ATE station was also one of the stations that registered the highest values of PM_{10} in Metropolitan Lima during the summer months; while, for that same period, CDM was one of the stations that recorded the lowest PM_{10} values during all seasons of the year [20].

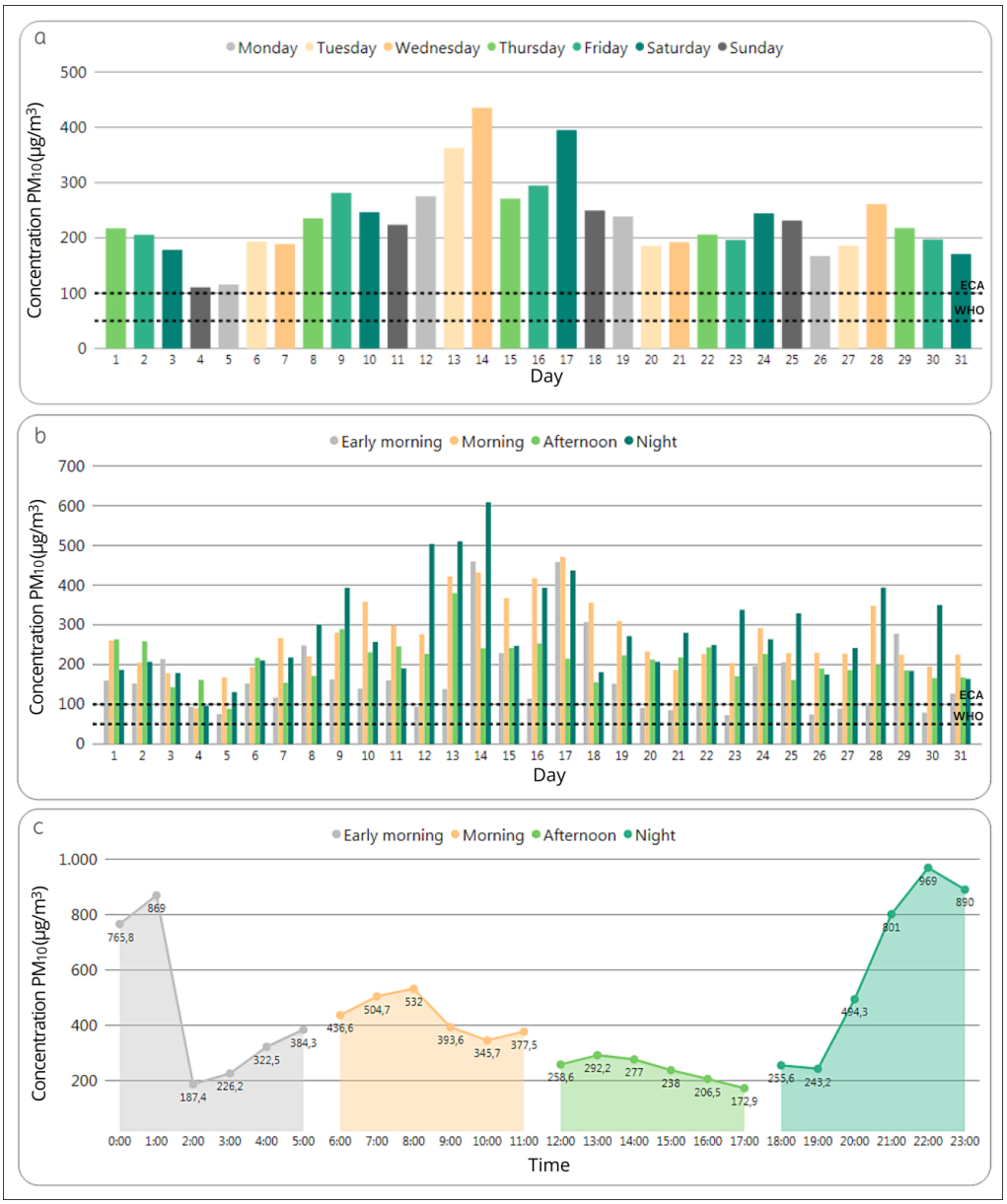

Figure 7. (**a**) Distribution of PM_{10} concentration by day in March 2018. (**b**) Distribution of PM_{10} concentration by episodes in March 2018. (**c**) Classification of PM_{10} concentration by episodes in 14 March 2018.

3.2. Detailed Analysis of the Station with the Highest Concentration of PM_{10}: HCH

According to the information in Figure 5, the highest concentrations of PM_{10} are observed at the HCH station. Therefore this station was chosen for a more detailed temporal analysis that is presented in this section. In this sense, Figure 6 shows the monthly, daily and hourly variation plots of the HCH station.

According to Figure 6a, the two months with the highest average concentrations during the study period for the HCH station were March and April of the year 2018, while the months of June, July and August in both years reported the lowest concentrations.

Since the highest values were reported in 2018, Figure 6b–d focus on this year. Based on Figure 6b it can be seen that the lowest concentrations are recorded on Sundays and Mondays, for most months. These results are comparable to those reported for Cali, Colombia [37] where PM_{10} concentrations increased from Monday to Saturday and decreased on Sunday when traffic flow in the city also decreased notably.

According to Figure 6d the hourly concentrations of PM_{10} behave in a varied way in each of the months of 2018. Only in the months of June, July and August a similar behavior can be observed that indicates lower values during the early morning, which rise until reaching a peak between 11:00 and 14:00.

In Figure 6c it ca be seen that March is the only month in which the concentrations reached values within the EPA categories "very unhealthy" and "hazardous", highlighting the 14th day, with a mean of 435.2 $\mu g/m^3$. Only 33.15% of the daily concentrations of PM_{10} registered during 2018 in the station HCH correspond to values lower than 100 $\mu g/m^3$, established by the ECA for daily means.

Figure 7 shows the variation of the PM_{10} concentration during the month of March, which, according to Figure 6 is the month that registered the highest values in 2018. Figure 7a shows an increase in the concentrations between the 9th and the 17th day of the month, with a maximum peak on the 14th. This may be due to the fact that the beginning of the school year, which represents an increase in traffic, was on March 12th. According to the Instituto Nacional de Estadística e Informática (National Statistical Institute of Peru), the national index of vehicular that flow through toll booths in March 2018 had an increase of 15.5% when compared to March in 2017 [38].

Figure 7b shows that the highest PM_{10} concentration during March was at night for three consecutive days: March 12th, 13th and 14th. On the other hand, in more than 80% of the days the early morning had the lowest PM_{10} values. Finally, Figure 7c shows that March 14 begins with very high PM_{10} concentrations between 0:00 and 1:00 in the morning. It should be noted that all hourly concentrations during March 14 exceed 170 $\mu g/m^3$, which indicates that all values are considered by the EPA in the categories unhealthy for sensitive people, unhealthy, very unhealthy or hazardous.

3.2.1. Temperature and Relative Humidity

In the summer of 2018 in the HCH station, the highest concentrations of PM_{10} and the highest temperatures were reached during the early morning, a low percentage of relative humidity was also recorded during that same time. Meanwhile, during the afternoon as the temperature decreases, the percentage of relative humidity increases and the levels of particulate matter decrease considerably.

As expected, in 2017 and 2018 the highest monthly temperature averages were recorded during the summer months and the lowest during the winter months.

PM_{10} concentrations could be influenced by the temperature, because higher temperatures increase photochemical activity, consequently, the decomposition of matter and an increase of PM_{10} values [39]. Also, the high temperatures allow the re-suspension of particulate material for long periods in the atmosphere [40]. On the other hand, relative humidity influences the sedimentation of particulate matter at ground level, making it more difficult to transport it in the air [37].

When analyzing the monthly averages of particulate material, it is observed that during the entire study period in the HCH station, concentrations higher than the ECA and WHO for daily averages were recorded. Figure 8a,b shows that the months with lower temperatures (June to October) favor the increase in relative humidity shown in Figure 8c,d, which could be related to the decrease in the levels of particulate material in the months under study.

The monthly averages of PM_{10} and temperature during 2018 in the HCH station show a strong positive Pearson correlation of 0.9, and a strong negative Pearson correlation between PM_{10} and relative humidity of -0.93 with . This does not imply a direct causality between the two meteorological variables and PM_{10} concentrations, because the increase in PM_{10} is also influenced by urban behavior events (e.g., industries and the automotive fleet).

3.2.2. Wind Speed and Wind Direction

Figure 9 shows the variation of wind speed and direction during the four episodes (parts of the day) in March 2018 for the HCH station. It is observed that the direction and

predominant speed of the wind varies during the day: in the early morning it takes the west direction with speeds of up to 6 m/s, and to a lesser percentage the southwest direction with speed up to 6 m/s; in the morning the predominant direction is the southwest, with winds up to 5.1 m/s; at afternoon, the southwest direction also prevails but with wind speeds up to 4.1 m/s. Finally, during the night, the predominant direction is the southwest with speed up to 5.1 m/s.

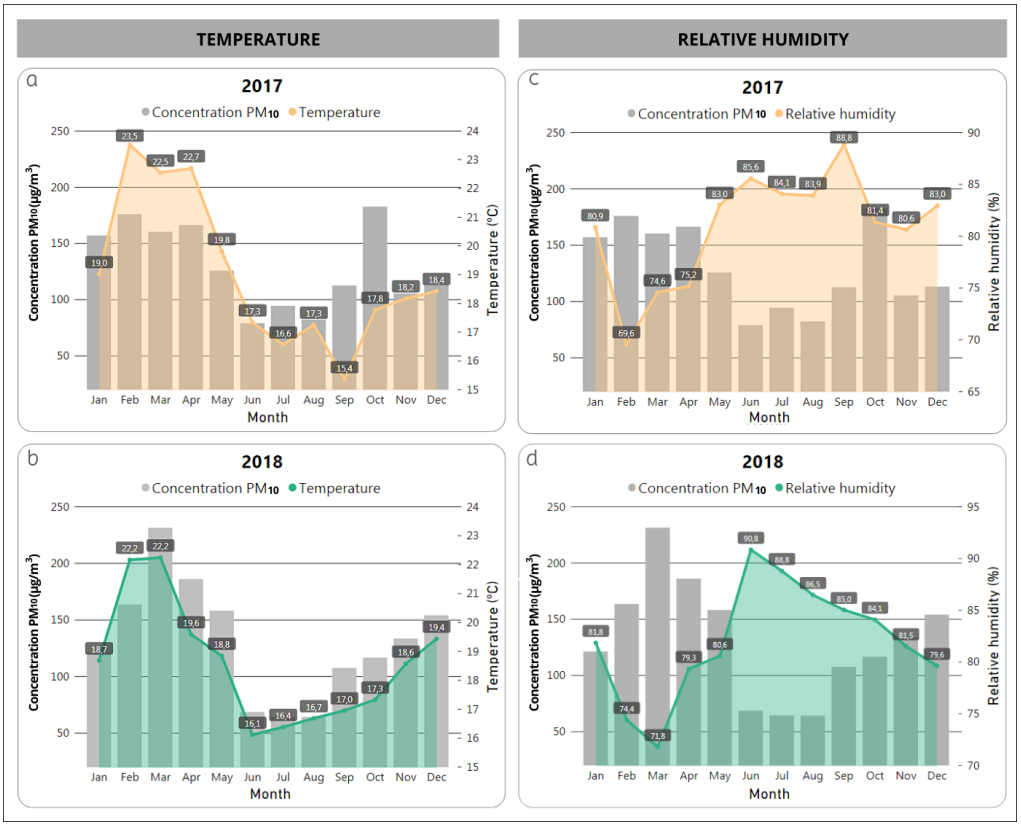

Figure 8. Monthly distribution of PM_{10} concentration and temperature in year (**a**) 2017 and (**b**) 2018 at the HCH station. Monthly distribution of PM_{10} concentration and relative humidity in year (**c**) 2017 and (**d**) 2018 at the HCH station.

3.3. Spatial Variation of PM_{10}

The intensity of the wind is greater in the fall from dawn until the night, being the lowest speed recorded between night and the next morning. Summer shows ups and downs in terms of wind intensities throughout the day, with a decrease in wind speed more noticeable during the afternoon and showing an increase during the early morning. During the winter, the greatest intensity is registered at noon and goes until sunset, with a decrease at night. Spring presents a similar behavior to the winter, with the greatest intensity being registered from noon until sunset, and when night falls, it diminishes until dawn.

3.3.1. Geolocated Wind Roses

The geolocated wind roses, reported in Figure 10, show the direction and speed of the wind. These graphs allow to complement the interpretation of the variation of PM_{10} in the Metropolitan Lima stations in 2018. It can be seen that the ATE, CDM, HCH and SMP stations present the same prevailing wind directions throughout the year, as well as similar ranges of wind speed in those directions. In contrast, for the CRB station, during the summer and fall the wind directions are more varied than in winter and spring. The CDM and CRB stations present the lowest concentrations of PM_{10}, with similar values throughout the year; the wind roses show a similar behavior in both stations. In the case of the CDM

station, the predominant direction of the wind is the southwest, which carries winds from the Pacific Ocean allowing the dispersion of pollutants in the air. For the CRB station, the highest wind speed is predominantly northwest during the four seasons of the year, which also come from the Pacific Ocean. On the other hand, the stations located east of Metropolitan Lima show the highest concentrations of PM_{10}, which may be related to the fact that both stations present a predominance of southwest winds from the districts located in the center of Metropolitan Lima throughout the year. An important factor to consider is that winds from the Pacific Ocean allow the dispersion of pollutants in the air; and in the case of the CRB station, the winds with the highest speed are predominantly northwest, i.e., from the Pacific Ocean during the whole year.

3.3.2. Heat Maps

Figure 11 shows the monthly heat maps for 2018. The stations located east of Metropolitan Lima (HCH and ATE) present the highest concentrations, in contrast, the stations located in the center (CDM) and north (SMP and CRB) have the lowest concentrations. This result coincide with the work of Delgado-Villanueva and Aguirre-Loayza [41], where the stations located to the east are classified as of very poor quality, while the stations in the center and south are described as of good quality.

Figure A1 shows that the stations with lower air quality are close to hills and high traffic roads such as the Ramiro Prialé highway, detailed in Figure A1a and central highway in Figure A1b. In addition, its surrounding streets are not paved, which influences the increase in the concentration of the levels of particulate matter [20]. In contrast, the stations in Figure A1c,d,f are those that register the least pollution and have green areas nearby; these influence the reduction of the temperature, thus increasing the relative humidity [42].

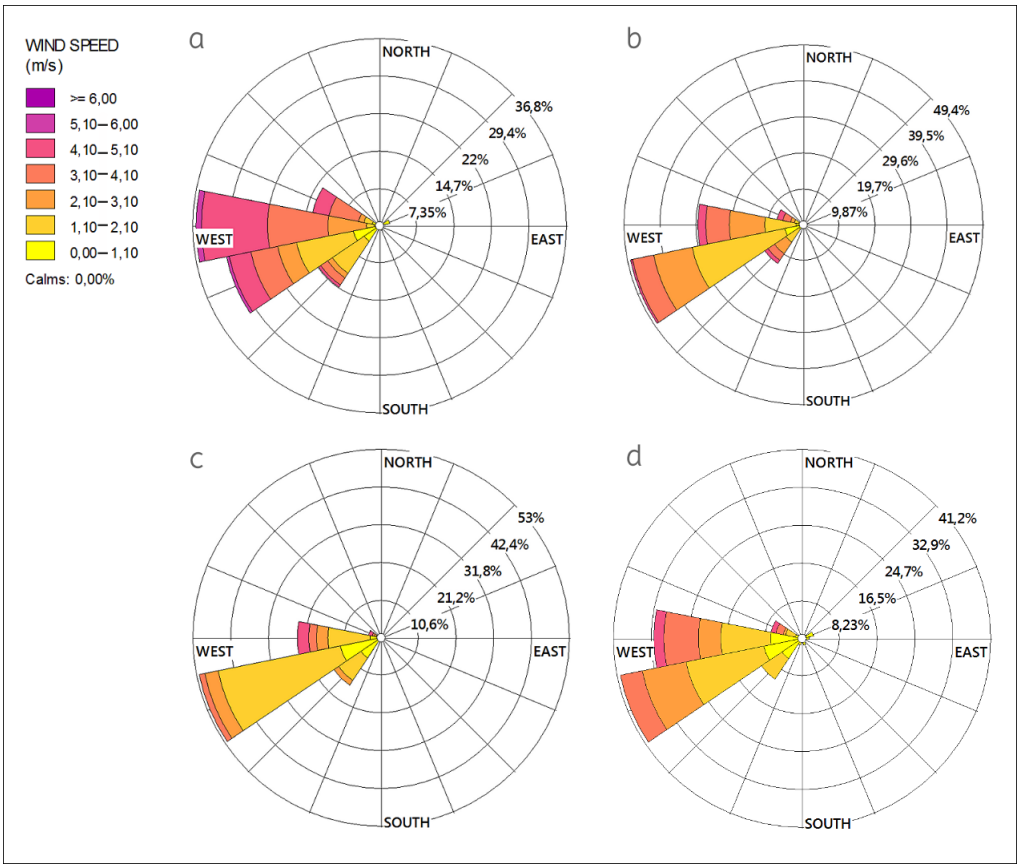

Figure 9. Wind roses by episodes:(**a**) early morning, (**b**) morning, (**c**) afternoon and (**d**) night in March 2018.

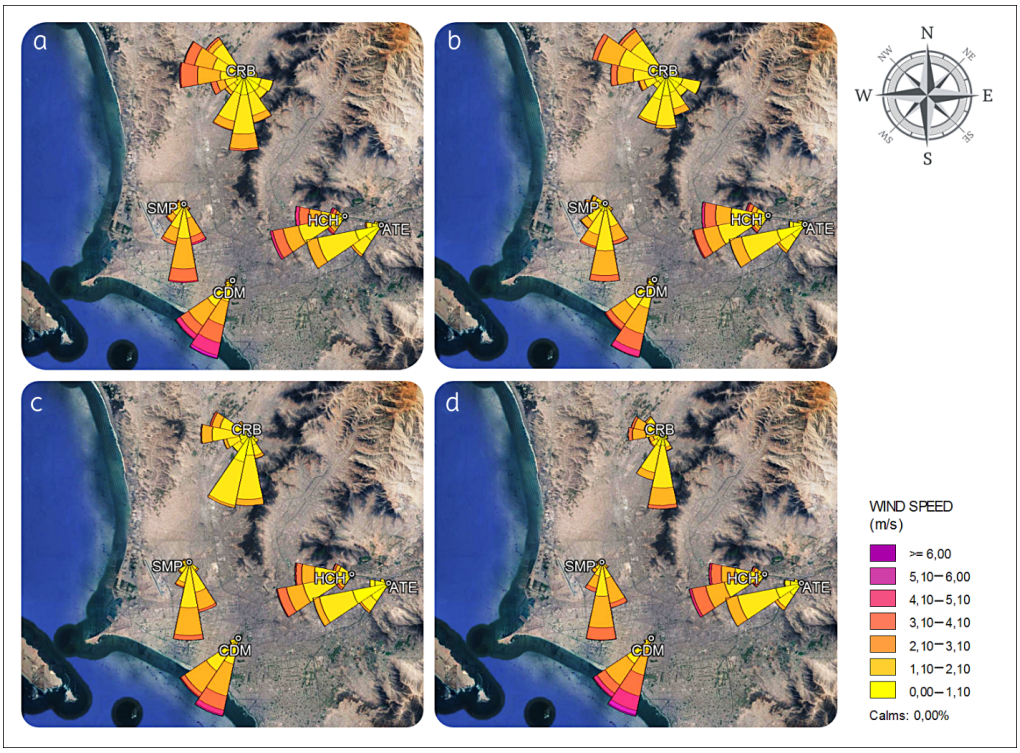

Figure 10. Geolocated wind roses of five air quality monitoring stations for each season in (**a**) summer, (**b**) fall, (**c**) winter, (**d**) spring of the year 2018.

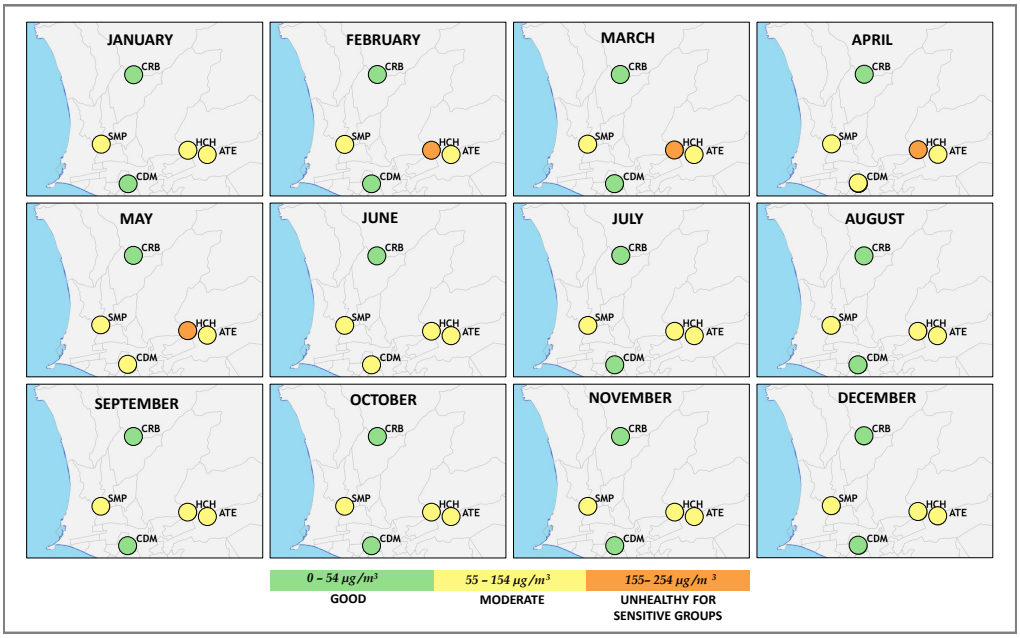

Figure 11. Heat maps with the monthly averages of PM_{10} concentration in 2018 for each monitoring station.

3.4. Hierarchical Clustering

Figure 12 shows the hierarchical tree (dendrogram) considering PM_{10} and the three meteorological variables, temperature, relative humidity and wind speed, for the five monitoring stations. Aiming at clustering time series, instead of data with independent observations and a standard Euclidean distance, we considered distances based on the dynamic time warping (DTW) algorithm, to obtain the dendrogram. In this way, a non-linear optimal alignment between each two time series is considered, instead of a point-

by-point distance. This dynamic correlational analysis between the different time series generated differentiable clusters, such as those made up of the same meteorological variable, which indicate that the distance between the meteorological data is minimal in the five stations of Metropolitan Lima. Likewise, it is observed that the wind speed is closer to the temperature than to the relative humidity. The CDM and CRB stations present the least distance in their PM_{10} behaviour in comparison to the other monitoring stations; both register the lowest concentrations of particulate matter, and they have a smaller distance with the relative humidity group than SMP. Finally, the two groups that show the least relationship between their PM_{10} behaviour and the other variables are HCH and ATE, both stations report the highest concentrations of particulate material throughout the study period.

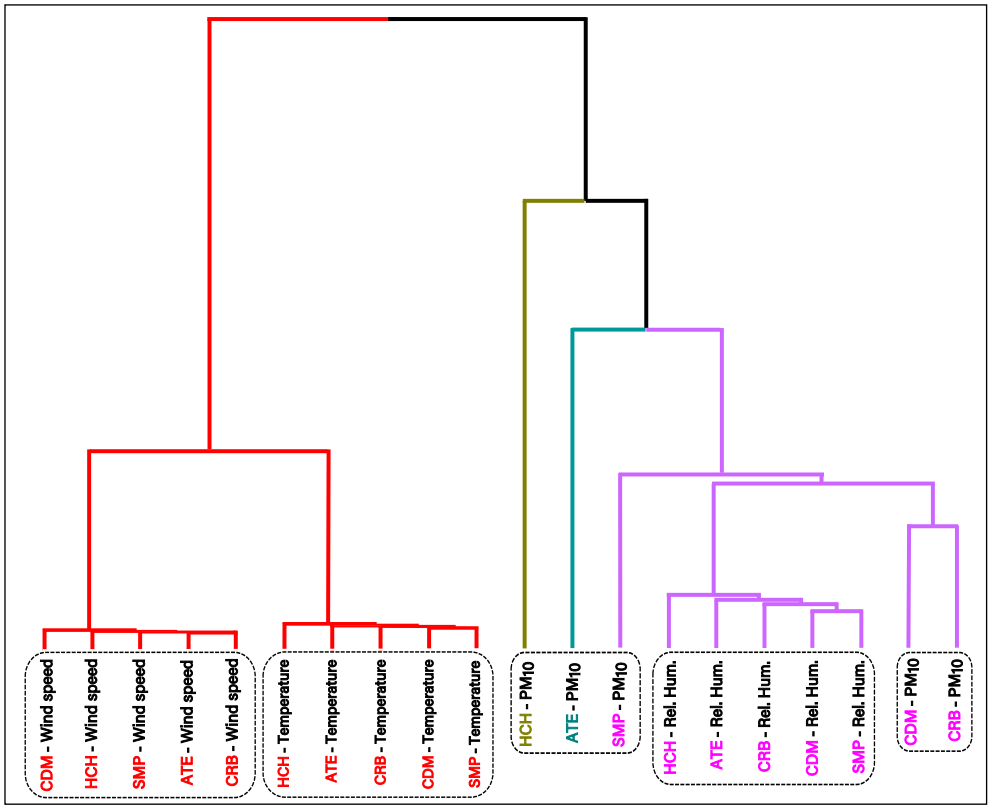

Figure 12. Plot hierarchical clustering dendrogram of meteorological variables and PM_{10} of the five monitoring stations as a result of the dynamic correlational analysis between the different time series. The generation of distinguishable clusters such as those made up of the same meteorological variable is observed.

4. Conclusions

The proposed method of a spatial-temporal visualization for the exploration of PM_{10} concentration data in Metropolitan Lima offers details in the analysis of the pollutant, allowing to identify the high concentrations registered in the study period. This would not have been possible if we had been used only plots with annual, monthly or seasonal averages.

According to the spatial distribution, the two stations located to the east of Metropolitan Lima (HCH and ATE) and one station located to the north (SMP) registered the highest average concentrations of PM_{10} during the study period. In the case of HCH and ATE, both stations are located near hills, high-traffic roads and unpaved streets. While the monitoring station located in the center (CDM) and one of the monitoring stations in the north (CRB) registered lower PM_{10} concentrations than the rest of the monitoring stations, both stations also have green areas around and predominant wind directions from the Pacific Ocean.

In relation to the temporal variation, the station with the highest average of biannual and annual PM_{10} was HCH. The highest PM_{10} concentrations were recorded in 2018, during the summer, highlighting the month of March with daily averages that reached up to 435 $\mu g/m^3$. In March 2018, the PM_{10} values increased from the 9th to the 17th, with a maximum peak on the 14th, which is likely to be related to the start of the school year on March 12. In 2018 only 33.15% of the daily PM_{10} concentrations were lower than the ECA (100 $\mu g/m^3$).

In 2018, the HCH station had a high positive correlation between the monthly average of PM_{10} and temperature, while the correlation was very low when applied on an hourly scale. This indicates that the relationship between the variables occurs with a time shift.

The closest relationships in the hierarchical cluster analysis occur between the meteorological variables indicating minimum distances between temperature, relative humidity and wind speed of the Metropolitan Lima stations; however, the behavior is different with the variable PM_{10} in the five stations. The meteorological variables do not indicate a causal relationship with respect to PM_{10}, but the concentrations of particulate matter would be more related to the urban characteristics of each district.

Previous studies about air quality data visualization in Metropolitan Lima [20,28] do not present a detailed step-by-step methodology to carry out a spatio-temporal visualization approach. The proposal developed in this paper, in addition to being specific in the approach to visualization, also presents adequate statistical techniques for the analysis of the time series related to environmental pollution. These analysis have a dynamic peculiarity in time, which allows us to better understand the different series in different monitoring stations in different time periods.

In terms of time series modelling and forecasting, further research should be conducted for this data set. Standard time series models such as exponential smoothing [43], as well as techniques such as the non-parametric singular spectrum analysis [44,45], relevant adaptations for non-negative time series (see, for instance, [46]), and hybrid methods [47–50], that have proven its usefulness in other areas and/or other data sets can be used to improve the understanding of this data and to conduct forecasting that can help with decision and policy making.

This approach will open bridges between the agencies in charge of monitoring air quality. Likewise, it will allow proposing policies in favor of caring for the environment, in order to mitigate air pollution. In addition, it opens the gate for future research in this field, extrapolating this proposal to other cities to facilitate analysis and improve decision-making at the local and national level.

Author Contributions: Conceptualization, A.A.E.-M. and J.D.C.-B.; methodology, J.L.L.-G., R.S. and P.C.R.; software, A.A.E.-M. and J.D.C.-B.; validation, J.L.L.-G., R.S. and P.C.R.; formal analysis, J.L.L.-G., R.S. and P.C.R.; investigation, A.A.E.-M., J.D.C.-B. and J.L.L.-G.; writing—original draft preparation,A.A.E.-M. and J.D.C.-B.; writing—review and editing, J.L.L.-G., R.S. and P.C.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially founded by the ANID scholarship.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: To the Servicio Nacional de Meteorología e Hidrología (SENAMHI) for providing the air quality data used in this study. Also to the ANID scholarship.

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

Abbreviations

The following abbreviations are used in this manuscript:

AQI	Air Quality Index
DIGESA	Dirección General de Salud Ambiental e Inocuidad Alimentaria
ECA	Estándar de Calidad Ambiental
EPA	Environmental Protection Agency
INEI	Instituto Nacional de Estadística e Informática
JICA	Japan International Cooperation Agency
LA	Latin America
LAC	Latin America and the Caribbean
SENAMHI	Servicio Nacional de Meteorología e Hidrología
WHO	World Health Organization

Appendix A

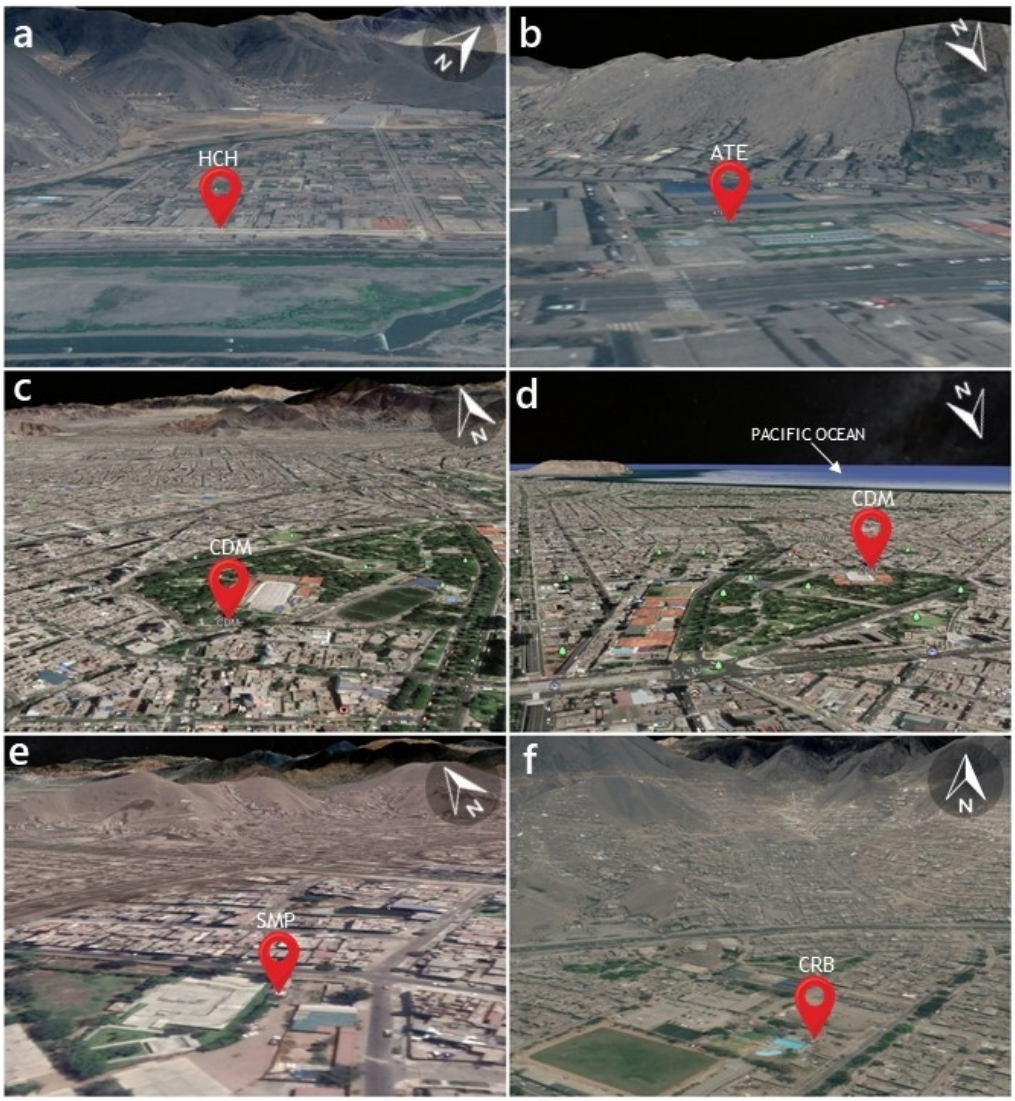

Figure A1. Satellite view of the monitoring stations (**a**) HCH, (**b**) ATE, (**c**) CDM, (**d**) CDM overlooking the Pacific Ocean, (**e**) SMP and (**f**) CRB.

References

1. Salas, R.; Bustos, A. Constructing a NARX model for the prediction of the PM_{10} air pollutant concentration. In Proceedings of the Encuentro Chileno de Computación, Jornada Chilena de Ciencias de la Computación, Valdivia, Chile, 7–12 November 2005.
2. Latorre, Á.M.L.R.; Tovar, M.H.T. Explotación minera y sus impactos ambientales y en salud. El caso de Potosí en Bogotá. *Saúde Em Debate* **2017**, *41*, 77–91. [CrossRef]
3. Chen, P. Visualization of real-time monitoring datagraphic of urban environmental quality. *EURASIP J. Image Video Process.* **2019**, *2019*, 42. [CrossRef]
4. Núñez Caraballo, V.; Rodríguez Rojas, R.; Gómez Camacho, L.; Herrera Moya, I.; Morales Pérez, M.C. Emisiones de dióxido de azufre a la atmósfera por fuentes fijas del MINAG y su influencia en la calidad del aire en la provincia de Villa Clara. *Cent. Agrícola* **2019**, *46*, 86–95.
5. Rojano, R.E.; Mendoza, Y.I.; Arregoces, H.; Restrepo, G.M. Dispersión de Contaminantes del Aire (PM_{10}, NO_2, CO, COV y HAP) emitidos desde una Estación Modular de Compresión, Tratamiento y Medición de Gas Natural. *Inf. Tecnológica* **2016**, *27*, 99–110. [CrossRef]
6. Nurgazy, M.; Zaslavsky, A.; Jayaraman, P.P.; Kubler, S.; Mitra, K.; Saguna, S. CAVisAP: Context-aware visualization of outdoor air pollution with IoT platforms. In Proceedings of the 2019 International Conference on High Performance Computing & Simulation (HPCS), Dublin, Ireland, 15–19 July 2019; pp. 84–91.
7. Park, Y.; Song, I.; Yi, J.; Yi, S.J.; Kim, S.Y. Web-Based Visualization of Scientific Research Findings: National-Scale Distribution of Air Pollution in South Korea. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2230. [CrossRef] [PubMed]
8. Tapia, V.; Carbajal, L.; Vásquez, V.; Espinoza, R.; Vásquez-Velásquez, C.; Steenland, K.; Gonzales, G.F. Traffic regulation and environmental pollution by particulate material (2.5 and 10), sulfur dioxide, and nitrogen dioxide in Metropolitan Lima, Peru. *Rev. Peru. De Med. Exp. Y Salud Publica* **2018**, *35*, 190. [CrossRef]
9. Zeng, Y.R.; Chang, Y.S.; Fang, Y.H. Data Visualization for Air Quality Analysis on Bigdata Platform. In Proceedings of the 2019 International Conference on System Science and Engineering (ICSSE), Dong Hoi, Vietnam, 20–21 July 2019; pp. 313–317.
10. EPA. *Midiendo los Avances: La Dimensión Ambiental de los OBJETIVOS de Desarrollo Sostenible en América Latina y el Caribe*; Technical Report; PNUMA Programa de las Naciones Unidas para el Medio Ambiente: Nairobi, Kenya, 2020; p. 109.
11. Sánchez, A.B.; Torres, B. *Una Recuperación Verde y Justa en América Latina y el Caribe: Una Perspectiva Desde el Mundo*; Technical Report; International Labour Organization: Geneva, Switzerland, 2020.
12. Riojas-Rodríguez, H.; da Silva, A.S.; Texcalac-Sangrador, J.L.; Moreno-Banda, G.L. Air pollution management and control in Latin America and the Caribbean: Implications for climate change. *Rev. Panam. De Salud Pública* **2016**, *40*, 150–159.
13. Air, I.Q. *2019 World Air Quality Report*; Technical Report; Index Quality Air; IQAir: Goldach, Switzerland, 2019; p. 25.
14. Poma, J.M.R. Desarrollo de un Modelo Dinámico para determinar la incidencia de los factores contaminantes del aire en la población de Lima Metropolitana. *Ind. Data* **2012**, *15*, 54–62. [CrossRef]
15. Valverde, J. Estudio de la calidad del aire afectada por la actividad industrial en la urb. Primavera-distrito de El Agustino. *Rev. Del Inst. De Investig. De La Fac. De Ing. Geológica, Minera, Metal. Y Geográfica* **2016**, *18*, 115–119.
16. Romero, Y.; Chicchon, N.; Duarte, F.; Noel, J.; Ratti, C.; Nyhan, M. Quantifying and spatial disaggregation of air pollution emissions from ground transportation in a developing country context: Case study for the Lima Metropolitan Area in Peru. *Sci. Total Environ.* **2020**, *698*, 134313. [CrossRef]
17. Aramayo, A. Percepción de la contaminación atmosférica en Lima: contraste entre la avenida Abancay y El Olivar de San Isidro. *Rev. Del Inst. De Investig. De La Fac. De Geol. Minas, Metal. Y Cienc. Geográficas* **2012**, *15*, 131–140.
18. Flores, E.C. Incidencia de la cuarentena por covid-19, en la calidad del aire (NO_2) de la ciudad de Lima. *Rev. Del Inst. De Investig. De La Fac. De Geol. Minas Metal. Y Cienc. Geográficas* **2020**, *23*, 65–71.
19. Suárez-Salas, L.; Álvarez Tolentino, D.; Bendezú, Y.; Pomalaya, J. Caracterización química del material particulado atmosférico del centro urbano de Huancayo, Perú. *Rev. De La Soc. Química Del Perú* **2017**, *83*, 187–199. [CrossRef]
20. Valdivia, S.A.P. Análisis temporal y espacial de la calidad del aire determinado por material particulado PM_{10} y PM2, 5 en Lima Metropolitana. *An. Científicos* **2016**, *77*, 273–283. [CrossRef]
21. JICA. *Encuesta de Recolección de Información Básica del Transporte Urbano en el Área Metropolitana de Lima y Callao. Informe Final*; Technical Report; Agencia de Cooperación Internacional del Japón: Tokyo, Japan 2013; p. 27.
22. Espinal, M.M. Transporte público de buses versus congestión y contaminación en Lima y Callao. *Economía* **2017**, *40*, 47–86. [CrossRef]
23. MINAM. *Diagnóstico de las Gestión de la Calidad Ambiental del Aire de Lima y Callao*; Technical Report; Ministerio del Ambiente: Lima, Peru, 2019.
24. Lu, W.; Ai, T.; Zhang, X.; He, Y. An interactive web mapping visualization of urban air quality monitoring data of China. *Atmosphere* **2017**, *8*, 148. [CrossRef]
25. Cat, J. Epistemology, aesthetics and pragmatics of scientific and other images: Visualization, representation and reasoning. In *Fuzzy Pictures as Philosophical Problem and Scientific Practice*; Springer: Berlin, Germany, 2017,; pp. 47–67.
26. San José, R.; Pérez, J.; González, R. Advances in 3D visualization of air quality data. *Usage Usability Utility 3D City Model.* **2012**, 02002. [CrossRef]
27. Li, H.; Fan, H.; Mao, F. A visualization approach to air pollution data exploration—A case study of air quality index ($PM_{2.5}$) in Beijing, China. *Atmosphere* **2016**, *7*, 35. [CrossRef]

28. Silva, J.; Rojas, J.; Norabuena, M.; Molina, C.; Toro, R.A.; Leiva-Guzmán, M.A. Particulate matter levels in a South American megacity: The metropolitan area of Lima-Callao, Peru. *Environ. Monit. Assess.* **2017**, *189*, 635. [CrossRef]
29. Reátegui-Romero, W.; Sánchez-Ccoyllo, O.R.; de Fatima Andrade, M.; Moya-Alvarez, A. PM2. 5 estimation with the WRF/Chem model, produced by vehicular flow in the lima metropolitan area. *Open J. Air Pollut.* **2018**, *7*, 215–243. [CrossRef]
30. Sánchez-Ccoyllo, O.R.; Ordoñez-Aquino, C.G.; Muñoz, Á.G.; Llacza, A.; Andrade, M.F.; Liu, Y.; Reátegui-Romero, W.; Brasseur, G. Modeling study of the particulate matter in lima with the WRF-Chem model: Case study of April 2016. *Int. J. Appl. Eng. Res. IJAER* **2018**, *13*, 10129. [CrossRef]
31. Buuren, S.V.; Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* **2010**, *45*, 1–68. [CrossRef]
32. Zhen, D.; Zhao, H.; Gu, F.; Ball, A. Phase-compensation-based dynamic time warping for fault diagnosis using the motor current signal. *Meas. Sci. Technol.* **2012**, *23*, 055601. [CrossRef]
33. Aghabozorgi, S.; Seyed Shirkhorshidi, A.; Ying Wah, T. Time-series clustering—A decade review. *Inf. Syst.* **2015**, *53*, 16–38. [CrossRef]
34. Salvador, S.; Chan, P. Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.* **2007**, *11*, 561–580. [CrossRef]
35. Hebbrecht, K.; Stuivenga, M.; Birkenhäger, T.; Morrens, M.; Fried, E.; Sabbe, B.; Giltay, E. Understanding personalized dynamics to inform precision medicine: A dynamic time warp analysis of 255 depressed inpatients. *BMC Med.* **2020**, *18*, 400. [CrossRef] [PubMed]
36. EPA. *National Ambient Air Quality Standards for Particulate Matter*; Technical Report; Environmental Protection Agency: Washington, DC, USA, 2013; Volume 78, p. 202.
37. Vinasco, J.P.S.; Nastar, T.d.R.C. Variación espacial y temporal de concentraciones de PM_{10} en el área urbana de Santiago de Cali, Colombia. *Ing. De Recur. Nat. Y Del Ambiente* **2013**, *12*, 129–141.
38. Tránsito de Vehículos a Nivel Nacional Aumentó 15.5%. 2018. Available online: https://n9.cl/b769 (accessed on 14 December 2020).
39. Jiménez, A.M.; Méndez, D.; Torrecilla, R.C. La concentración de partículas en el aire: Análisis estadístico de la relación espacial entre medidas de superficie y del sensor MODIS para dos tipos de tiempo en la Comunidad de Madrid. *Investig. Geográficas* **2020**, *73*, 189–209. [CrossRef]
40. Guzmán, A.A.E.; Tzuc, O.M.; Pantí, I.B.; Trujeque, J.R.; Quintana, I.V.P.; Bassam, A. Artificial neural network modeling of PM_{10} and $PM_{2.5}$ in a tropical climate region: San Francisco de Campeche, Mexico. *Química Nova* **2017**, *40*, 1025–1034.
41. Delgado-Villanueva, A.; Aguirre-Loayza, A. Modelamiento y evaluación del nivel de calidad del aire mediante el análisis de grey clustering, estudio de caso Lima metropolitana. *Tecnia* **2020**, *30*, 114–120. [CrossRef]
42. Zegarra-Peña, R.; Andrade-Tenesaca, S.; Parra-Ullauri, M.; Mejía-Coronel, D.; Rodas-Espinoza, C. Análisis espacial de PM_{10} en el aire y su composición de metales con relación a factores ambientales alrededor de centros de educación preescolar en Cuenca. *Maskana* **2020**, *11*, 57–68. [CrossRef]
43. Sulandari, W.; Suhartono.; Subanar.; Rodrigues, P.C. Exponential Smoothing on Modeling and Forecasting Multiple Seasonal Time Series: An Overview. *Fluct. Noise Lett.* **2021**, 2130003. [CrossRef]
44. Rodrigues, P.C.; Mahmoudvand, R. The benefits of multivariate singular spectrum analysis over the univariate version. *J. Frankl. Inst.* **2018**, *355*, 544–564. [CrossRef]
45. Rodrigues, P.C.; Pimentel, J.; Messala, P.; Kazemi, M. The decomposition and forecasting of mutual investment funds using singular spectrum analysis. *Entropy* **2020**, *22*, 83. [CrossRef] [PubMed]
46. Awe, O.O.; Mahmoudvand, R.; Rodrigues, P.C. Non-Negative Time Series Reconstruction via Singular Spectrum Analysis: A Case Study of Precipitation Dynamics in Nigeria. *Fluct. Noise Lett.* **2020**, *19*, 2050045. [CrossRef]
47. Sulandari, W.; Subanar; Lee, M.H.; Rodrigues, P.C. Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks. *Energy* **2020**, *190*, 116408. [CrossRef]
48. Sulandari, W.; Subanar, S.; Suhartono, S.; Utami, H.; Lee, M.H.; Rodrigues, P.C. SSA-based hybrid forecasting models and applications. *Bull. Electr. Eng. Inform.* **2020**, *9*, 2178–2188. [CrossRef]
49. Sulandari, W.; Subanar, S.; Lee, M.H.; Rodrigues, P.C. Time series forecasting using singular spectrum analysis, fuzzy systems and neural networks. *MethodsX* **2020**, *7*, 101015. [CrossRef]
50. Rodrigues, P.C.; Awe, O.O.; Sousa, J.; Mahmoudvand, R. Modelling the Behaviour of Currency Exchange Rates with Singular Spectrum Analysis and Artificial Neural Networks. *Stats* **2020**, *3*, 137–157. [CrossRef]