



OPEN

## Statistical modeling approach for PM<sub>10</sub> prediction before and during confinement by COVID-19 in South Lima, Perú

Rita Jaqueline Cabello-Torres<sup>1✉</sup>, Manuel Angel Ponce Estela<sup>2</sup>, Odón Sánchez-Ccoyllo<sup>3</sup>, Edison Alessandro Romero-Cabello<sup>4</sup>, Fausto Fernando García Ávila<sup>5</sup>, Carlos Alberto Castañeda-Olivera<sup>1</sup>, Lorgio Valdiviezo-Gonzales<sup>6</sup>, Carlos Enrique Quispe Eulogio<sup>7</sup>, Alex Rubén Huamán De La Cruz<sup>8</sup> & Javier Linkolk López-Gonzales<sup>9</sup>

A total of 188,859 meteorological-PM<sub>10</sub> data validated before (2019) and during the COVID-19 pandemic (2020) were used. In order to predict PM<sub>10</sub> in two districts of South Lima in Peru, hourly, daily, monthly and seasonal variations of the data were analyzed. Principal Component Analysis (PCA) and linear/nonlinear modeling were applied. The results showed the highest annual average PM<sub>10</sub> for San Juan de Miraflores (SJM) (PM<sub>10</sub>-SJM: 78.7 µg/m<sup>3</sup>) and the lowest in Santiago de Surco (SS) (PM<sub>10</sub>-SS: 40.2 µg/m<sup>3</sup>). The PCA showed the influence of relative humidity (RH)-atmospheric pressure (AP)-temperature (T)/dew point (DP)-wind speed (WS)-wind direction (WD) combinations. Cool months with higher humidity and atmospheric instability decreased PM<sub>10</sub> values in SJM and warm months increased it, favored by thermal inversion (TI). Dust resuspension, vehicular transport and stationary sources contributed more PM<sub>10</sub> at peak times in the morning and evening. The Multiple linear regression (MLR) showed the best correlation ( $r = 0.6166$ ), followed by the three-dimensional model LogAP-LogWD-LogPM<sub>10</sub> ( $r = 0.5753$ ); the RMSE-MLR (12.92) exceeded that found in the 3D models (RMSE < 0.3) and the NSE-MLR criterion (0.3804) was acceptable. PM<sub>10</sub> prediction was modeled using the algorithmic approach in any scenario to optimize urban management decisions in times of pandemic.

Air is vital and its quality is measured in the most strategic places in cities<sup>1,2</sup>. Particulate matter (PM) with an aerodynamic diameter < 10µm represents a global health problem<sup>3</sup> with short or long term effects<sup>4,5</sup>. The fixation of this pollutant in the upper respiratory system depends on the inhalation flow rate<sup>6</sup>, causing cardiovascular diseases<sup>7</sup> and mortality due to lung cancer<sup>8</sup>. Megacities are the most affected due to industrialization, energy consumption and high vehicular flow, generating social costs<sup>9</sup>.

The World Health Organization (WHO)<sup>10</sup> has established PM<sub>10</sub> concentration threshold values for 24-h (50 µg/m<sup>3</sup>) and annual (20 µg/m<sup>3</sup>) average measurements, which are used as a reference in air quality monitoring programs. Meteorological monitoring also contributes to measuring the effects of meteorological variables on air quality and the environment<sup>11</sup>. Relative humidity (RH), temperature (T), atmospheric pressure (AP), wind speed (WS) and wind direction (WD) influence both the distribution and concentration of PM<sub>10</sub><sup>12-16</sup>. This information is used in predictive models of particulate matter to optimize the control of atmospheric emissions<sup>17</sup> and to manage sustainable cities<sup>18</sup>.

The city of Lima in Peru, is approaching ten million inhabitants and the appearance of the COVID-19 pandemic in March 2020 led to the closure of land borders<sup>19</sup> and suspension of international air transport, leading

<sup>1</sup>Universidad César Vallejo, Escuela de Ingeniería Ambiental, Lima, Peru. <sup>2</sup>Dirección General de Salud Ambiental, Lima, Peru. <sup>3</sup>Universidad Nacional Tecnológica de Lima Sur, Lima, Peru. <sup>4</sup>Universidad Nacional Agraria La Molina, Escuela de Ingeniería Ambiental, Lima, Peru. <sup>5</sup>Universidad de Cuenca, Facultad de Ciencias Químicas, Grupo RISKEN, Cuenca, Ecuador. <sup>6</sup>Universidad Tecnológica del Perú, Facultad de Ingeniería Mecánica e Industrial, Lima, Peru. <sup>7</sup>Facultad de Ciencias de la Salud, Universidad Peruana Los Andes, Huancayo, Peru. <sup>8</sup>E.P. de Ingeniería Ambiental, Universidad Nacional Intercultural de la Selva Central Juan Santos Atahualpa, La Merced, Peru. <sup>9</sup>Facultad de Ingeniería y Arquitectura, Universidad Peruana Unión, Lima, Peru. ✉email: rcabello@ucv.edu.pe

to an improvement in air quality<sup>3,20,21</sup> until the start of the economic reactivation in May 2020 that activated the industrial sector and vehicular transport (approximately 22,794,136 daily trips in 2019)<sup>22,23</sup>.

The *Ministerio del Ambiente* (MINAM) updated the air quality standards (AQS)<sup>24</sup> and established the air quality index (AQI) for PM<sub>10</sub> called Índice de Calidad del Aire (INCA)<sup>25</sup>, but there is no evidence of PM<sub>10</sub> prediction models in southern Lima to help improve air quality management and minimize risks to human health. The *Dirección General de Salud Ambiental* (DIGESA) of the *Ministerio de Salud* (MINSA) maintains a monitoring network for relative humidity (RH), atmospheric pressure (AP), solar radiation (SR), wind speed (WS), wind direction (WD) and PM<sub>10</sub> as part of the air quality health surveillance program in Lima and Callao. This network includes two stations in southern Lima: *Santiago de Surco* (SS), an urbanized and commercial sector; and *San Juan de Miraflores* (SJM), a less paved and dusty sector. Two areas of Lima that are complete opposites in terms of environmental management characteristics. In addition, SJM borders the “Panamericana Sur” highway, one of the most important roads in the country and close to cement factories. Both districts (SS and SJM) have a total of 900,000 inhabitants<sup>26</sup>.

In this context, statistical models are suitable to describe the complex relationship between PM<sub>10</sub> and meteorological variables, predicting its behavior especially in urban areas. This research applies statistical approaches to the prediction of PM<sub>10</sub> before and during the pandemic caused by COVID-19 in Lima. Moreover, there are a limited number of studies in Lima, which is one of the cities with the highest pollution levels in South America<sup>27–29</sup>. Predicting air quality with high accuracy can be problematic, but these tools are becoming increasingly important because they provide comprehensive information to prevent critical pollution episodes and reduce human exposure to this pollutants<sup>30,31</sup>.

The objective of this research is to contribute to the environmental management of air quality through the application of a simple and effective statistical modeling of air quality related to PM<sub>10</sub> concentration levels, based on meteorological parameters and with scientific support that allows authorities to optimize decision making in the control of air pollution and risks to human health. The contributions of the research are summarized below:

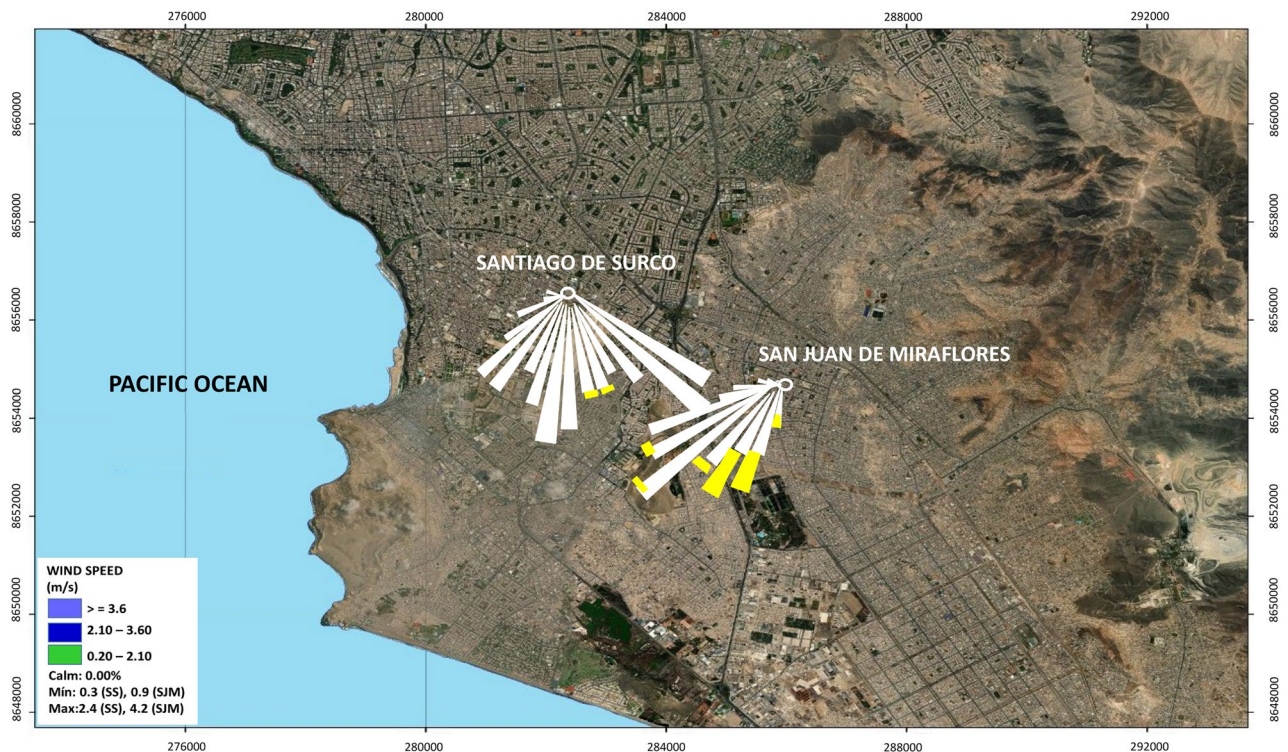
- Implementation of statistical modeling for time series data during 2019–2020, at two meteorological and air quality monitoring locations in southern Lima.
- The three-dimensional PM<sub>10</sub> forecasting model based on time series, being the first time this analysis is applied in South Lima. In addition, the principal component analysis (PCA) to evaluate the effect of meteorological variables on the behavior of PM<sub>10</sub>.

The rest of the paper is structured as follows: “[Literature review](#)”, shows the different studies that precede this research. Then, “[Materials and methods](#)” that describes the methodology developed based on statistical modeling approaches. Also, “[Results and discussion](#)” that presents the main findings of this research compared to other studies. Finally, “[Limitations](#)” and “[Conclusions](#)” that provides the main conclusions, together with some recommendations for future research.

## Literature review

The review of studies with applications for modeling PM<sub>10</sub> as a function of meteorological variables did not produce information in southern Lima (SJM and SS stations), but there is some background information for other districts belonging to the city of Lima. For example, air quality was evaluated at 3 points in North Lima, using the gray grouping method<sup>32</sup>. In addition, the dependence of particulate matter in air on meteorological parameters was studied in the town of *Zárate* in Lima<sup>33</sup>. Also, PM concentrations were analyzed one month before the pandemic (February 2020) and at the beginning of the pandemic (March–April 2020)<sup>21</sup>. In another investigation, a multivariate regression model based on internodal correlations ranging from 0.31 to 0.49 was implemented to analyze air quality at other stations in Lima, using a low-cost sensor with IIoT technology<sup>18</sup>. Similarly, the Weather Research and Forecasting-Chem (WRF-Chem) model was applied to develop an operational forecast system for air quality in the Metropolitan Area of Lima and Callao (MALC)<sup>12</sup>. Air quality for PM<sub>10</sub> was also evaluated in other districts of Lima through the application of artificial neural networks (ANN), observing certain difficulties for its prediction in stations with critical pollution episodes<sup>34</sup>.

Applications of PM<sub>10</sub> models on meteorological variables are related to the use of algorithms and functional relationships<sup>35</sup>. In China, meteorological conditions were related to interannual PM variations through correlations using robust ANN prediction models<sup>36</sup>. In Iran, meteorological factors were related to the PM<sub>10</sub>/PM<sub>2.5</sub> ratio using the AirQ+ model, calculating positive correlations between relative humidity (RH) and temperature (T), and negative correlations with precipitation and wind speed (WS)<sup>37</sup>. However, one of the most commonly used models is the multiple linear regression (MLR) model<sup>16,38</sup>. For example, in Bulgaria, MLR was applied relating AP-RH-T-T-PM<sub>10</sub>, reporting moderate correlations and indicating that more factors affecting PM<sub>10</sub> concentration should be included<sup>15</sup>. In China, PM<sub>10</sub> and PM<sub>2.5</sub> were related to meteorological variables by linear and single-factor exponential regressions, and determined robust models of negative relationship for WS-PM<sub>10</sub> and positive relationships for T-PM<sub>10</sub> in warm seasons, but not in cool seasons<sup>14</sup>. Also, ANN and MLR models were compared for meteorological variable-PM<sub>10</sub> combinations without obtaining significant correlations ( $p > 0.05$ ) for RH<sup>39</sup>. Another study combined the dew point (DP)-T variables on an integral Moivre-Laplace model to predict PM<sub>10</sub> and PM<sub>2.5</sub><sup>40</sup>. Similarly, the T-WS-AP-RH-PM<sub>10</sub> factors were used, showing limitations in the single factor models, and instead applied 3D graph fits to obtain better prediction results<sup>13</sup>. In the present research, the methodology proved to be simple and robust and was based on the statistical modeling approach to predict PM<sub>10</sub> concentration and contribute to generate new tools for air quality management, aerosol control and prevention of risks associated with COVID-19.



**Figure 1.** Location map and wind rose for SS and SJM, period 2019–2020. Map created in Arcgis software version 10.4.1 (<https://desktop.arcgis.com/es/arcmap/10.4/get-started/setup/arcgis-desktop-quick-start-guide.htm>).

## Materials and methods

**Area of study and dataset.** Lima, the capital of Peru, has a flat morphology formed by the valleys of the *Chillón*, *Rímac* and *Lurín* rivers, and meteorological variables are influenced by the relief of the Andes mountain range, the cold Humboldt current and the South Pacific anticyclone (SPA), which generates microclimates in the city<sup>41</sup> and variations in the altitudes of the TI base that influence PM<sub>10</sub> dispersion.

The first monitoring station is located at a MINSA facility in the SS district (12°8'43.47"S 76°59'50.07"W), and the second is located at the *Hospital María Auxiliadora* in the SJM district (12°9'41.33"S 76°57'32.36"W). The monitoring generates continuous information and the publication of real-time data for PM<sub>10</sub> (µg/m<sup>3</sup>), temperature-T (°C), dew point-DP (°C), atmospheric pressure-AP (hPa), solar radiation-SR (W/m<sup>2</sup>), wind speed-WS (m/s) and wind direction-WD (degrees, °). The map of Lima south-Peru (see Fig. 1) shows the monitoring stations evaluated in the period 2019–2020. This map was prepared using the *Arcgis 10.4.1 software*, with the shapefiles from Peru, including the Pacific Ocean.

From Fig. 1, both monitoring stations correspond to coastal and surrounding urban areas in southern Lima. The district of *San Juan de Miraflores* is more desert and sandy, with prevailing coastal winds, intense vehicular traffic, dust resuspension<sup>42</sup> and close to limestone, clay and gypsum quarries. The predominant wind rose presented South (S), South South West (SSW) and South East (SE) directions, with intensities between 0.9 and 2.5 m/s, with a punctual anomalous value in the spring (4.2 m/s) for SJM and between 0.3 to 2.4 m/s for SS.

**Equipment and materials.** The calibrated automatic equipment (zero air equipment and argon gas dilution equipment) were installed between 1.5 and 10 m from the ground and connected to 220 volts, avoiding solid barriers 10 meters around to ensure adequate air flow. The Campbell Scientific weather station was calibrated to measure each meteorological variable. Data obtained from January 1, 2019 to December 31, 2020 were processed by real-time telemetry using AirMetReport software. Glass fiber of 25 mm diameter was used, placed daily inside the low volume test equipment (PM<sub>10</sub> beta gauge) with automatic analysis of PM<sub>10</sub> concentration in µg/m<sup>3</sup>. Monitoring was performed according to the Protocol for Air Quality Monitoring and Data Management<sup>43</sup>. The data were validated by DIGESA and are published on the institutional web page.

**Statistical procedure.** A total of 188,859 valid hourly data (T, DP, AP, SR, RH, WD, WS and PM<sub>10</sub>) generated between 2019 and 2020 were used. For 2019 only 83.5% of the data were used, while for 2020 only 51.13% were used. The 2019 data were used to develop the models and their calibration, and the 2020 data were used to validate or evaluate the models obtained in 2019. The hourly scale data were used for plotting temporal variability and statistical analysis. For its part, the wind rose analysis was elaborated using Wrplot View V.8.0.2 software. The input data were used in Origin-Pro 8.0 software, generating the following:



- Analysis of the correlation between meteorological factors and PM<sub>10</sub>.
- Application of PCA and dimensionality reduction of the interrelated variables by linear transformation of the input vectors from high to low dimension, and generation of uncorrelated components by reducing the number of predictor variables through the correlation matrix (M):

$$|M - \lambda I| \tag{1}$$

where  $\lambda$  is an eigenvalue and  $I$  is the identity matrix,  $\lambda$  multiplied by a nonzero eigenvector  $E$  generates the correspondence  $C_e$

$$C_e = \lambda E \tag{2}$$

Thus, the  $j$ -th variance of the  $j$ -th principal component PC is given as:

$$\text{Variance} = \frac{\kappa_j}{\sum_n \lambda_n} \tag{3}$$

The calculated principal components (PC) generate a maximum  $\lambda$  of linear combination of variables with the highest data variability, transforming the original set to the orthogonal one by multiplying the eigenvectors<sup>44</sup>.

- PM<sub>10</sub> modeling using a single meteorological variable.
- PM<sub>10</sub> modeling using MLR, defined by:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \varepsilon_i, \quad i = 1, 2, \dots, n \tag{4}$$

where  $Y_i$  the dependent variable;  $\beta_0, \beta_1, \dots, \beta_k$  are  $k + 1$  constant parameters;  $X_1, \dots, X_k$  are  $k$  independent variables;  $\varepsilon$  are the independent errors identically and Gaussian distributed and  $n$  is the number of observations<sup>39</sup>.

- Modeling of PM<sub>10</sub> using the combination of 2 meteorological variables by 3D surface fitting, which is an extension of the ordinary nonlinear fitting for both XYZ and matrix data. The method consisted of converting all data to Log10, assigning the meteorological variables (independent) to the X, Y axes and the PM<sub>10</sub> (dependent) to the Z axis. The worksheet was converted to matrix using the grid method and random parameters of 24 columns by 24 rows. Different models based on the Levenberg-Marquardt algorithm (damped least squares) were tested taking into account the metrics defined in this item, generating the following functions: Extreme Cum: Non-linear Extreme Value Cumulative Function

$$z = z_0 + B * \exp \left\{ - \exp \left\{ \frac{C - x}{D} \right\} \right\} + E * \exp \left\{ - \exp \left\{ \frac{F - y}{G} \right\} \right\} + H \exp \left\{ - \exp \left\{ \frac{C - x}{D} \right\} - \exp \left\{ \frac{F - y}{G} \right\} \right\} \tag{5}$$

Voigt2DMod: The voigt surface with volume as parameter, donde:  $z_0, A, x_c, w_1, y_c, w_2, m_u$  are constant parameters.

$$z = z_0 + A \left[ m_u \frac{4}{\pi^2} * \frac{w_1}{4(x - x_c)^2 + w_1^2} * \frac{w_2}{4(y - y_c)^2 + w_2^2} + (1 - m_u) \frac{4Ln2}{\pi w_1 w_2} e^{\frac{-4Ln2}{w_1^2} (x - x_c)^2 - \frac{-4Ln2}{w_2^2} (y - y_c)^2} \right] \tag{6}$$

Poly2D: Two-dimensional polynomial function

$$z = z_0 + ax + by + cx^2 + dy^2 + fxy \tag{7}$$

For these tests, 95% reliability, correlation coefficient (r) and coefficient of determination (R<sup>2</sup>) were considered<sup>45,46</sup>.

**Performance metrics.** Analysis of prediction performance involves calculating the errors between the observed and predicted values. Four statistical metrics were used to compare the performance of the models:

1. Pearson correlation coefficient (r):

$$r = \frac{\sum_{i=1}^n (Y_o^i - \bar{Y}_o)(Y_m^i - \bar{Y}_m)}{\sqrt{\sum_{i=1}^n (Y_o^i - \bar{Y}_o)^2} \cdot \sqrt{\sum_{i=1}^n (Y_m^i - \bar{Y}_m)^2}} \tag{8}$$

2. Coefficient of determination (R<sup>2</sup>):

$$R^2 = \frac{[\sum_{i=1}^n (Y_o^i - \bar{Y}_o)(Y_m^i - \bar{Y}_m)]^2}{\sum_{i=1}^n (Y_o^i - \bar{Y}_o)^2 * \sum_{i=1}^n (Y_m^i - \bar{Y}_m)^2} \tag{9}$$

3. Root Mean Squared Error (RMSE):

Rating	Care	Values range	Interval ( $\mu\text{g}/\text{m}^3$ )	Colour
Good	Satisfactory air quality	0–50	0–75	Green
Moderate	Acceptable air quality	51–100	76–150	Yellow
Poor	Health problems are experienced	101–VUEC	151–250	Orange
TVSC*	Causes health effects	> VUEC	> 250	Red

**Table 1.** Air quality index values. \*Threshold value of state of care.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (Y_m^i - Y_o^i)^2}{n}} \quad (10)$$

4. Nash-Sutcliffe Efficiency (NSE):

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (Y_m^i - Y_o^i)^2}{\sum_{i=1}^n (Y_o^i - \bar{Y}_o)^2} - \alpha < \text{DC} \leq 1.0 \quad (11)$$

where  $Y_o^i$ ,  $Y_m^i$ , stand for model predicted and target values, respectively,  $\bar{Y}_o$ ,  $\bar{Y}_m$ , are their mean values and  $n$  represents the number of observations.

**Air quality index (AQI) and air quality standards (AQS).** In Peru, the AQI  $\text{PM}_{10}$  for 24 h ( $100 \mu\text{g}/\text{m}^3$ ) should not be exceeded more than 7 times a year and the annual arithmetic mean should not exceed  $50 \mu\text{g}/\text{m}^3$ . The AQI  $\text{PM}_{10}$  was calculated from the 24-h AQS  $\text{PM}_{10}$  and the alert threshold value ( $150 \mu\text{g}/\text{m}^3$ ) according to the following expression:

$$I(\text{PM}_{10}) = \frac{[\text{PM}_{10} \mu\text{g}/\text{m}^3] 100 \mu\text{g}/\text{m}^3}{150 \mu\text{g}/\text{m}^3} \quad (12)$$

where  $I(\text{PM}_{10})$  expresses the calculated AQI  $\text{PM}_{10}$ , and the value inside the parenthesis is the observed  $\text{PM}_{10}$ . Table 1 shows the national AQI classification (INCA).

## Results and discussion

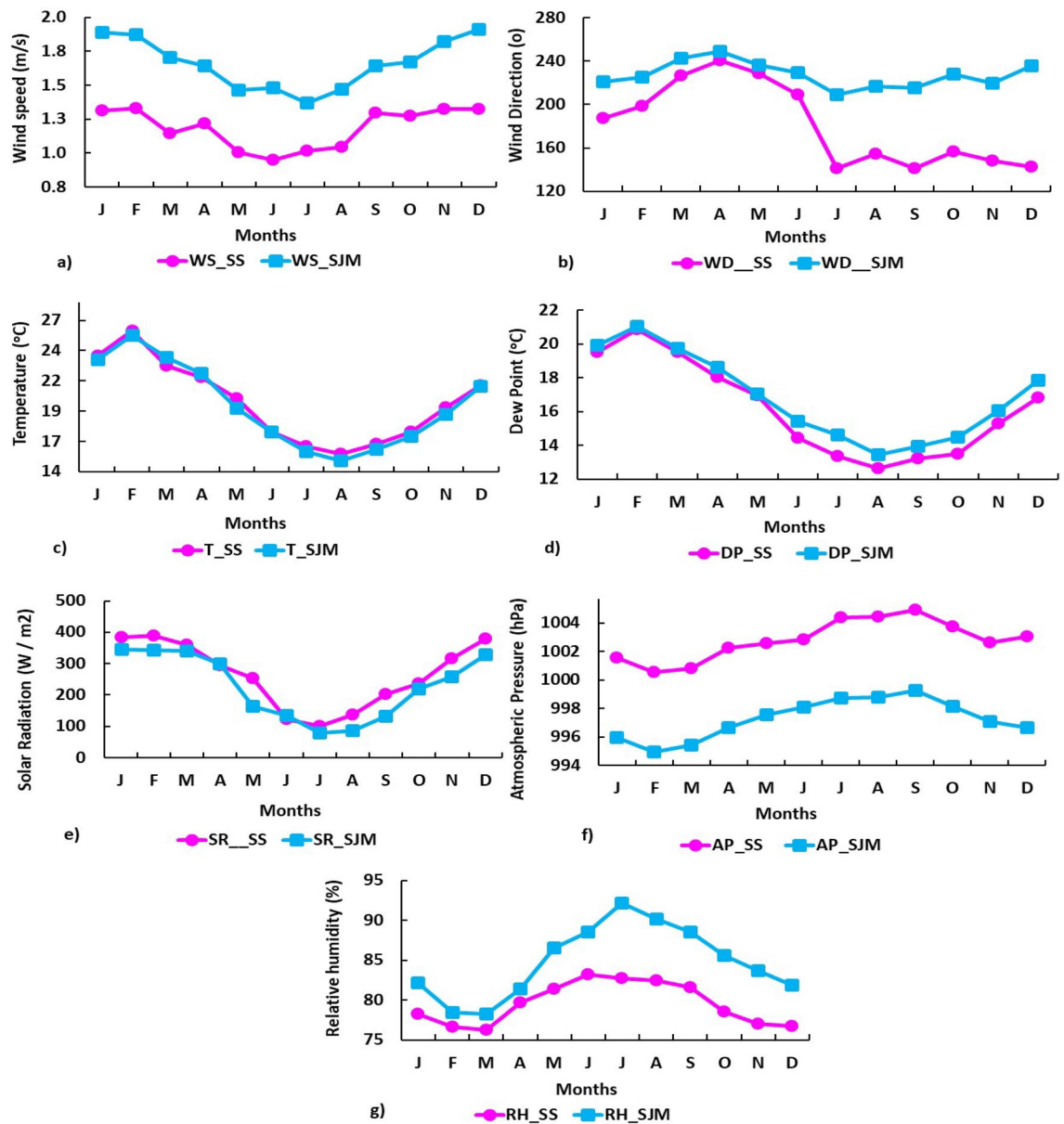
**Meteorological variations and seasonal correlations.** Figure 2 shows the average monthly distribution of meteorological variables in SS and SJM during 2019. The monthly averages of meteorological variables evidenced that:

- T, DP and SR variables are higher in the austral summer (January, February and March), while RH and AP have low monthly averages due to the proximity of the sun to the earth.
- In winter, especially in July and August, there is a decrease in SR and T (including DP), and the values of AP and RH increase.

WS varied discretely from 1.06 m/s in January (SS) a 1.93 m/s in December (SJM) (Fig. 2a), with wind patterns in a south-easterly direction for SS and south-westerly for SJM, especially in April (Fig. 2b). Temperature ranged between  $15.11^\circ\text{C}$  in winter (August-SJM) and  $25.56^\circ\text{C}$  in summer (February-SS) (ver Fig. 2c), and DP fluctuated between  $12.61^\circ\text{C}$  (August-SS) and  $21.18^\circ\text{C}$  (February-SJM) (see Fig. 2d).

The annual means of T and DP in both districts showed that  $T_{SS} > DP_{SS}$  ( $3.65^\circ\text{C}$ ) and  $T_{SJM} > DP_{SJM}$  ( $2.59^\circ\text{C}$ ), but SS was warmer than SJM:  $T_{SS} > T_{SJM}$  ( $0.26^\circ\text{C}$ ) and  $DP_{SJM} > DP_{SS}$  ( $0.68^\circ\text{C}$ ). Likewise, SR was higher in summer ( $394.36 \text{ W}/\text{m}^2$ , February-SS) and lower in winter ( $80.26 \text{ W}/\text{m}^2$ , July-SJM) (Fig. 2e), with annual means  $SR_{SS} > SR_{SJM}$  ( $34.1 \text{ W}/\text{m}^2$ ) indicating that the Peruvian central coastal strip presents greater annual variations of solar energy received over the surface<sup>47</sup>. In contrast, AP (see Fig. 2f) and RH (see Fig. 2g) showed an inverse behavior and particular patterns ( $RH_{SS} < RH_{SJM}$  and  $AP_{SS} > AP_{SJM}$ ), with mean monthly values of RH ranging between 69.8% (Summer-March, SS) and 91.94% (Winter-July, SJM), while AP fluctuated between 994.99 hPa (February, SJM) and 1004.73 hPa (September, SS), with maximum peaks in the winter months. Figure 3 shows the variability of meteorological factors in southern Lima, representing the temporal evolution of  $\text{PM}_{10}$  in *Santiago de Surco* and *San Juan de Miraflores* during 2019.

**Hourly, daily, weekly and monthly variation of  $\text{PM}_{10}$  concentration.** Figure 3a,b, show lower  $\text{PM}_{10}$  hourly averages in the early morning (1:00 a.m. and 5:00 a.m.), associated with higher humidities and weaker winds that kept aerosols suspended. As the hours passed, T increased and RH decreased, concentrating  $\text{PM}_{10}$  in the air column, especially at peak hours in the morning (9:00 a.m. and 12:00 p.m.) and at night (6:00 p.m. and 9:00 p.m.). In SS,  $\text{PM}_{10}$  at peak hours followed the seasonal order: spring ( $60.38 \mu\text{g}/\text{m}^3$ ) > autumn ( $49.64 \mu\text{g}/\text{m}^3$ ) > winter ( $44.73 \mu\text{g}/\text{m}^3$ ), > summer ( $40.92 \mu\text{g}/\text{m}^3$ ) and in SJM: summer ( $117.9 \mu\text{g}/\text{m}^3$ ) > autumn ( $107.8 \mu\text{g}/\text{m}^3$ ) > spring ( $102.1 \mu\text{g}/\text{m}^3$ ) > winter ( $91.3 \mu\text{g}/\text{m}^3$ ). The intense activity of the Lima-Callao vehicle fleet<sup>48</sup>, dust resuspension and industrial activities generate this effect.

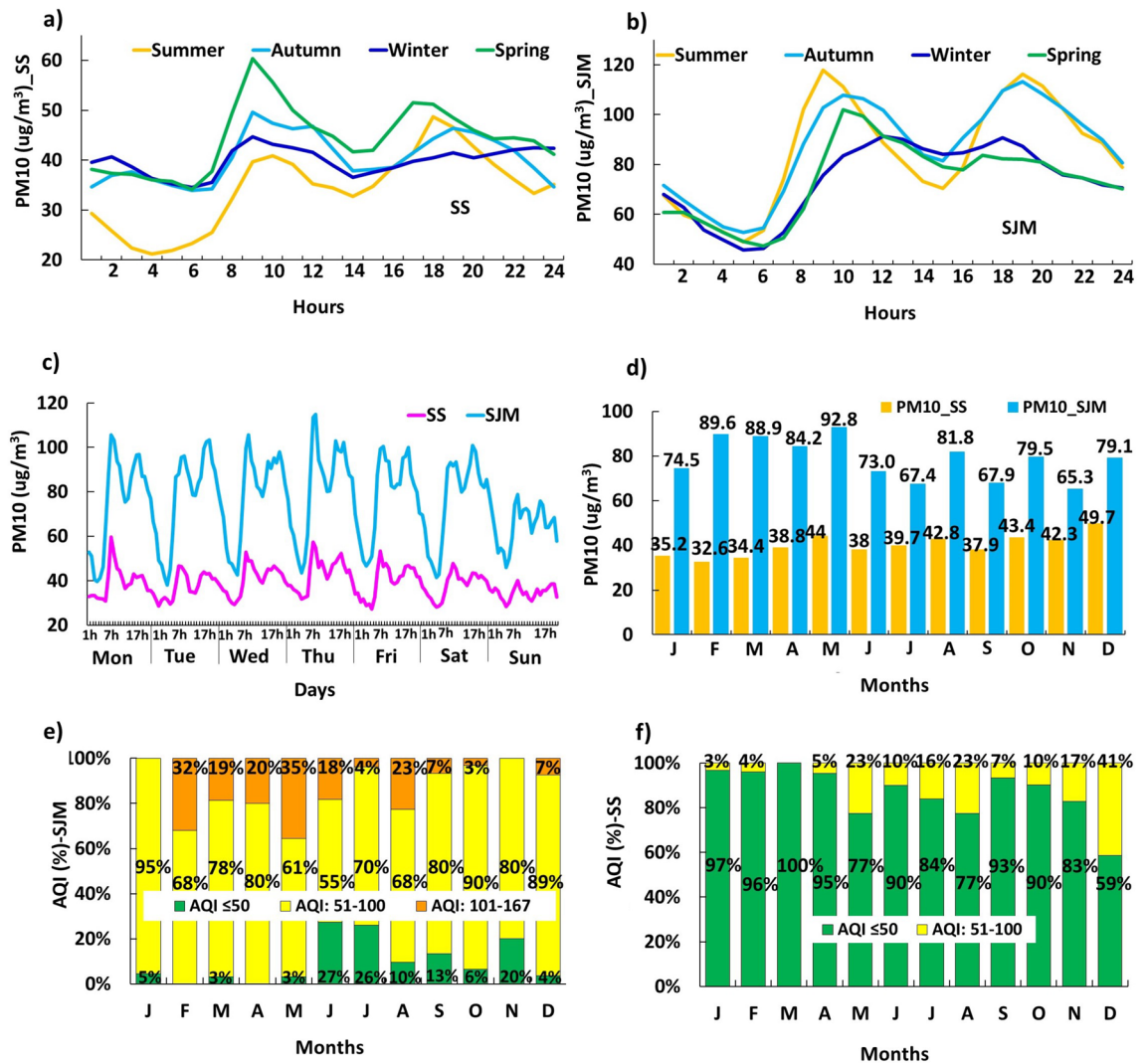


**Figure 2.** Distribution of the monthly mean values of the meteorological variables at stations SS and SJM, between January and December 2019, Lima-Peru: (a) WS (m/s), (b) WD (degrees), (c) T (°C), (d) DP (°C), (e) SR (W/m<sup>2</sup>), (f) AP (hPa) and (g) RH (%).

Likewise, Fig. 3c shows the daily averages of PM<sub>10</sub>, in SS and SJM. Lower values were recorded on Saturdays and Sundays (SS: 29 μg/m<sup>3</sup> and SJM : 54.58 μg/m<sup>3</sup>), due to formal and student work breaks, and decreased activity of vehicular mobile sources<sup>48</sup>. For the other days of the week, the values were higher (Wednesday in SS: 46 μg/m<sup>3</sup>, in spring; and Monday in SJM: 97.94 μg/m<sup>3</sup>, in autumn).

On the other hand, the monthly mean values for PM<sub>10</sub> (Fig. 3d) ranged from 32.6 μg/m<sup>3</sup> (February-SS) to 92.8 μg/m<sup>3</sup> (May-SJM). Post hoc tests produced for SJM a significant difference ( $p = 0.0$ ) between concentrations recorded during the warm austral summer-autumn months (PM<sub>10</sub>: 89.6 to 92.8 μg/m<sup>3</sup>) and cool winter-spring months (PM<sub>10</sub>: 73 to 79.1 μg/m<sup>3</sup>). The decrease in TI base altitude in the coastal summer-autumn coincided with the highest PM<sub>10</sub> concentrations.

**Air quality indexes.** In 2019, the AQIs values in South Lima (Fig. 3e,f) were 49.3 % good quality (PM<sub>10</sub> < 76 μg/m<sup>3</sup>) > 44.1% moderate quality (76 – 150 μg/m<sup>3</sup>) > 6.5% poor quality (101-250 μg/m<sup>3</sup> threshold state of care). In addition, differences were found between SS and SJM AQIs. The AQI<sub>SJM</sub> values were 77% of moderate quality > 14% of poor quality > 8.8% of good quality, while the AQI<sub>SS</sub> values were 86% of good quality > 14% of moderate quality. On the other hand, 13% of hourly PM<sub>10SS</sub> values exceeded the WHO reference value (50 μg/m<sup>3</sup>) but their annual mean 40 μg/m<sup>3</sup> did not exceed the WHO annual reference value (50 μg/m<sup>3</sup>); while for SJM, 100% of hourly data exceeded the WHO reference. Likewise, 15% of hourly PM<sub>10SJM</sub> values exceeded

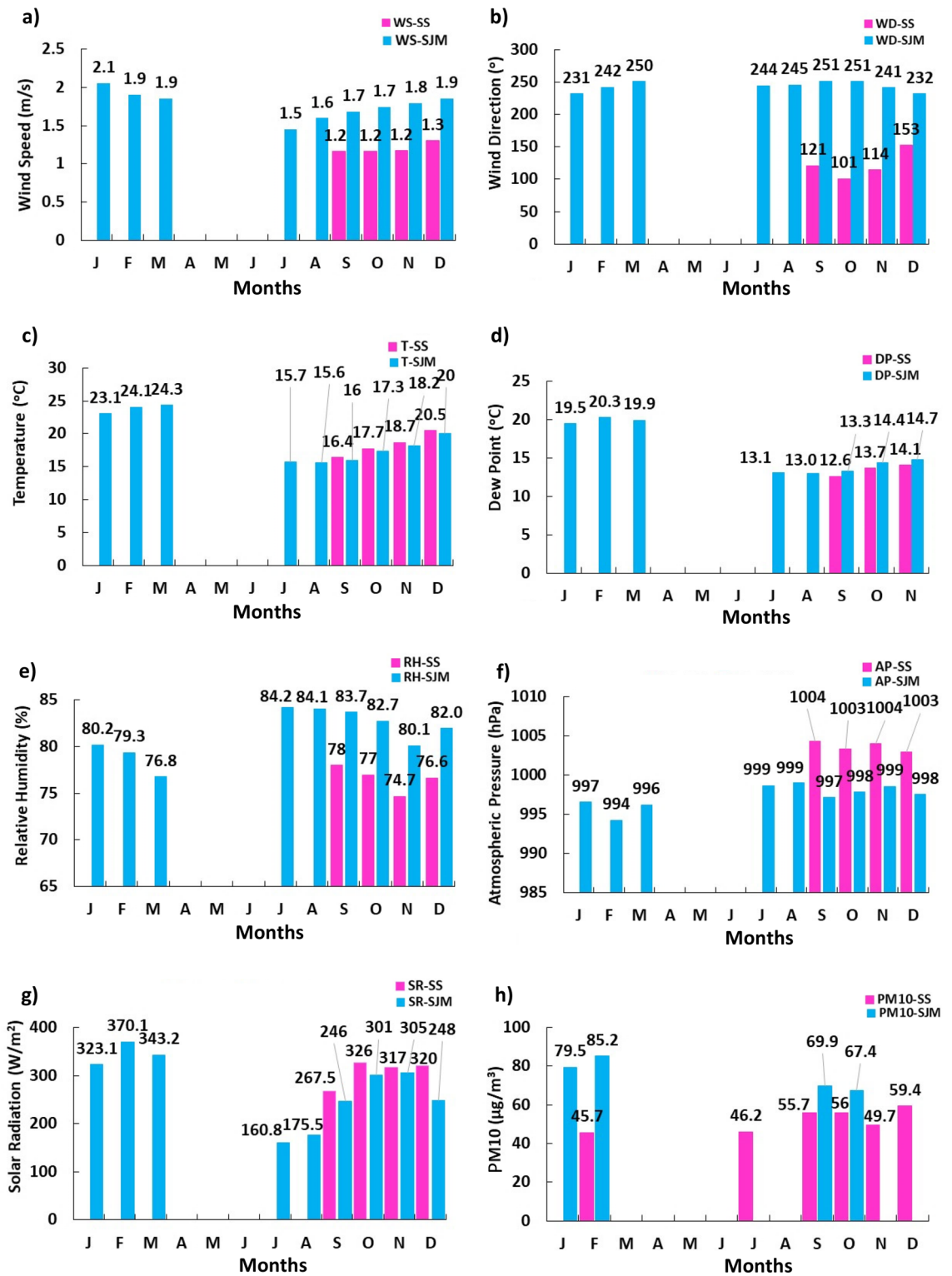


**Figure 3.** Temporal variation of PM<sub>10</sub>: (a) Hourly mean in SS; (b) Hourly mean in SJM; (c) Daily behavior of PM<sub>10</sub> throughout the week; (d) Monthly mean; (e) Percentage of AQI values in SJM; (f) Percentage of AQI values in SS. Period 2019.

the current national AQS (100 μg/m<sup>3</sup>) and the annual mean (78.7 μg/m<sup>3</sup>) exceeded the annual AQS (50 μg/m<sup>3</sup>). The fact is that both locations present a potential risk to respiratory and cardiovascular diseases, especially SJM due to the daily exposure of people to higher PM<sub>10</sub> levels.

**Meteorological and PM<sub>10</sub> variations in 2020 during the COVID-19 pandemic.** Due to the pandemic, annual monitoring of meteorological variables in SJM was suspended between April and June and PM<sub>10</sub> was monitored in January, February, September and October. In SS, meteorological monitoring was continuous and PM<sub>10</sub> was monitored in February and July-December. On the other hand, the increase in monthly average RH in South Lima was notorious, as SR<sub>SJM2020</sub> (274.74 W/m<sup>2</sup>) > SR<sub>SJM2019</sub> (227.37 W/m<sup>2</sup>) and SR<sub>SS2020</sub> (307.68 W/m<sup>2</sup>) > SR<sub>SS2019</sub> (258.68 W/m<sup>2</sup>). The wind pattern for SS had dominant east-southeast direction and for SJM southwest direction. In 2020, the comparison of PM<sub>10</sub> concentrations for SJM corresponding to the months of each year showed no significant differences, while for SS a 25% increase PM<sub>10</sub> concentration was observed. The pandemic did not produce reduced PM<sub>10</sub> levels in 2020 as expected due to the operability of major emission sources during that time<sup>49</sup>. The monitoring results are shown in Fig. 4.

**Correlations between meteorological variables and PM<sub>10</sub>.** Table 2, shows the correlations determined for the meteorological variables and PM<sub>10</sub> data observed in 2019. Significant statistical values (p < 0.05), with strong and moderate magnitudes, are shown in bold: r<sub>T-DP</sub> (0.95788) > r<sub>T-SR</sub> (0.72471) > r<sub>RH-SR</sub> (-0.66936) > r<sub>RH-T</sub> (-0.6484) > r<sub>DP-SR</sub> (0.61907). Regarding PM<sub>10</sub>, the order was moderate and weak: r<sub>WD-PM<sub>10</sub></sub> (0.48192) > r<sub>WS-PM<sub>10</sub></sub> (0.40526) > r<sub>AP-PM<sub>10</sub></sub> (-0.39443) > r<sub>RH-PM<sub>10</sub></sub> (0.18348) > r<sub>DP-PM<sub>10</sub></sub> (0.13796). Likewise, Fig. 5 shows the fluctuations of each meteorological vari-



**Figure 4.** Distribution of the monthly mean values of the meteorological variables at the SS and SJM stations, between January and December 2020: (a) WS (m/s), (b) WD (degrees), (c) T (°C), (d) DP (°C), (e) AP (hPa) and (f) RH (%), (g) SR (W/m<sup>2</sup>) and (h) PM<sub>10</sub> (µg/m<sup>3</sup>).

able and PM<sub>10</sub> concentrations. While, Fig. 6 presents the single variable regressions in Cartesian coordinates for southern Lima in the period 2019.

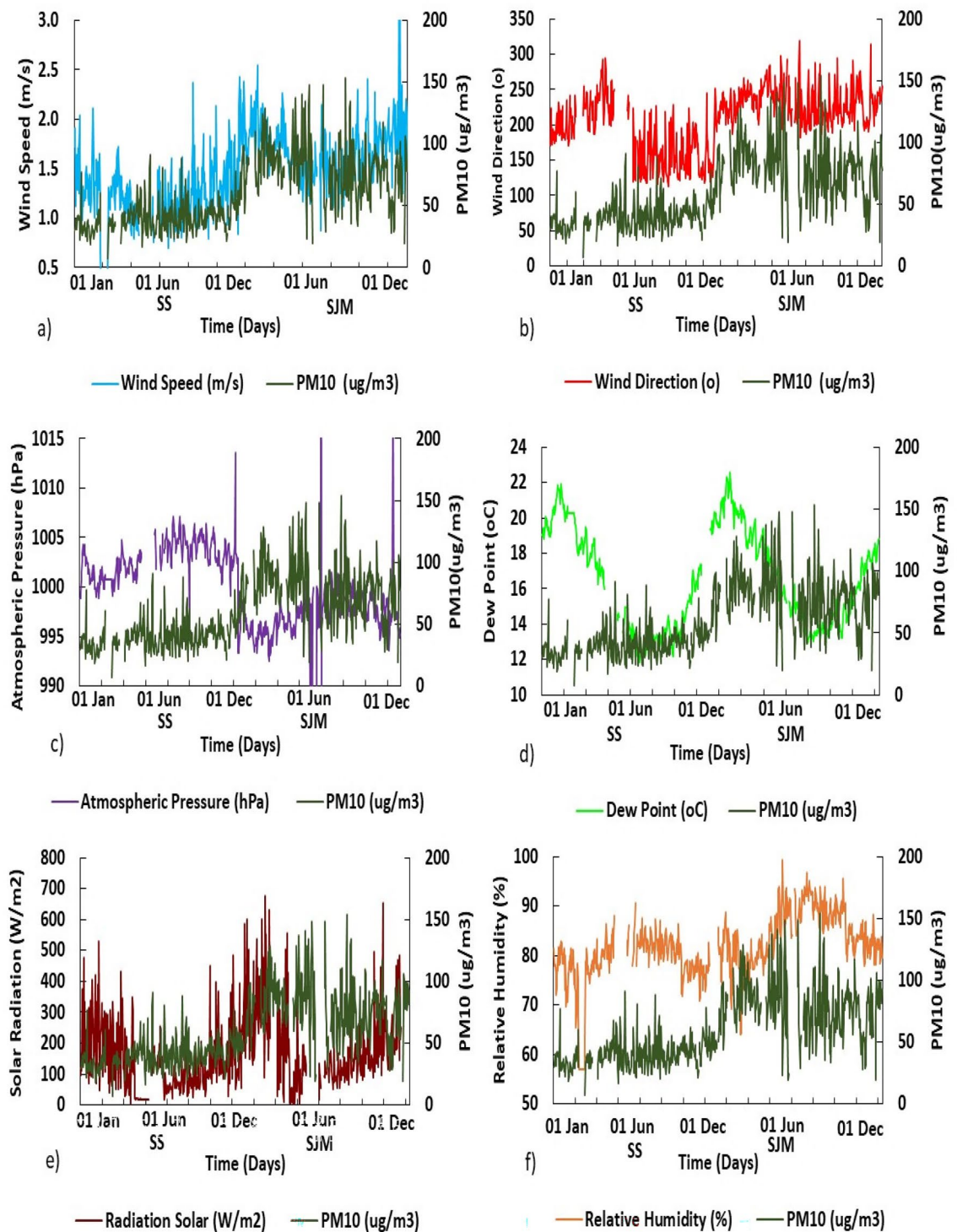
The wind speed and direction presented in Fig. 5a, shows the average daily fluctuation of WS in southern Lima, with a U-shaped distribution for SS and SJM, and Fig. 5b shows the fluctuation for WD. Seasonally,



	RH	T	DP	AP	SR	WS	WD	PM <sub>10</sub>
<b>Summer</b>								
RH	1	-0.67751	0.056	-0.33923	-0.26431	0.66721	0.2313	0.02371
T	-0.67751	1	0.64416	-0.05572	0.34623	0.04159	-0.15641	0.08682
DP	0.056	0.64416	1	-0.27044	0.15988	0.08832	0.02631	0.07581
AP	-0.33923	-0.05572	-0.27044	1	0.13959	-0.59284	-0.59514	-0.74682
SR	-0.26431	0.34623	0.15988	0.13959	1	0.09111	-0.11085	-0.21155
WS	0.66721	0.04159	0.08832	-0.59284	0.09111	1	0.01221	0.55296
WD	0.2313	-0.15641	0.02631	-0.59514	-0.11085	0.01221	1	0.55989
PM <sub>10</sub>	0.02371	0.08682	0.07581	-0.74682	-0.21155	0.55296	0.55989	1
<b>Autumn</b>								
RH	1	-0.80644	-0.49657	-0.3199	-0.67986	-0.0763	0.21508	0.22361
T	-0.80644	1	0.90638	0.09043	0.81919	0.30554	-0.13793	-0.04332
DP	-0.49657	0.90638	1	0.05445	0.74073	0.39667	-0.00969	0.10543
AP	-0.3199	0.09043	0.05445	1	0.15268	-0.42613	-0.07857	-0.72994
SR	-0.67986	0.81919	0.74073	0.15268	1	0.43466	-0.14084	0.01694
WS	-0.0763	0.30554	0.39667	-0.42613	0.43466	1	-0.06862	0.56274
WD	0.21508	-0.13793	-0.00969	-0.07857	-0.14084	-0.06862	1	0.08837
PM <sub>10</sub>	0.22361	-0.04332	0.10543	-0.72994	0.01694	0.56274	0.08837	1
<b>Winter</b>								
RH	1	-0.41628	0.58515	-0.16319	-0.55647	0.39848	0.54395	0.35943
T	-0.41628	1	0.49301	0.12102	0.38446	0.04495	-0.18351	-0.06938
DP	0.58515	0.49301	1	-0.0305	-0.20124	0.42113	0.35397	0.2957
AP	-0.16319	0.12102	-0.0305	1	0.05873	-0.07583	-0.24061	-0.23791
SR	-0.55647	0.38446	-0.20124	0.05873	1	0.02819	-0.05578	0.14091
WS	0.39848	0.04495	0.42113	-0.07583	0.02819	1	0.23279	0.29453
WD	0.54395	-0.18351	0.35397	-0.24061	-0.05578	0.23279	1	0.5149
PM <sub>10</sub>	0.35943	-0.06938	0.2957	-0.23791	0.14091	0.29453	0.5149	1
<b>Spring</b>								
RH	1	-0.48193	0.04568	-0.47784	-0.53996	0.26821	0.54183	0.39192
T	-0.48193	1	0.85274	-0.03936	0.43651	0.21912	-0.0587	-0.04819
DP	0.04568	0.85274	1	-0.33073	0.1734	0.41321	0.2611	0.184
AP	-0.47784	-0.03936	-0.33073	1	0.05839	-0.25586	-0.71338	-0.54587
SR	-0.53996	0.43651	0.1734	0.05839	1	0.23968	0.00886	0.14659
WS	0.26821	0.21912	0.41321	-0.25586	0.23968	1	0.40479	0.33795
WD	0.54183	-0.0587	0.2611	-0.71338	0.00886	0.40479	1	0.72592
PM <sub>10</sub>	0.39192	-0.04819	0.184	-0.54587	0.14659	0.33795	0.72592	1
<b>Annual</b>								
RH	1	-0.64912	-0.44149	-0.08189	-0.67403	0.23093	0.23457	0.18833
T	-0.64912	1	0.95778	-0.21475	0.7253	0.28424	0.22679	0.05294
DP	-0.44149	0.95778	1	-0.257	0.6178	0.20475	0.36127	0.12721
AP	-0.08189	-0.21475	-0.257	1	-0.08977	-0.24426	-0.37568	-0.38676
SR	-0.67403	0.7253	0.6178	-0.08977	1	0.34519	0.10545	0.04713
WS	0.23093	0.28424	0.20475	-0.24426	0.34519	1	0.17999	0.39432
WD	0.23457	0.22679	0.36127	-0.37568	0.10545	0.17999	1	0.47482
PM <sub>10</sub>	0.18833	0.05294	0.12721	-0.38676	0.04713	0.39432	0.47482	1

**Table 2.** Pearson correlations of the meteorological variables and the annual and seasonal PM<sub>10</sub>.

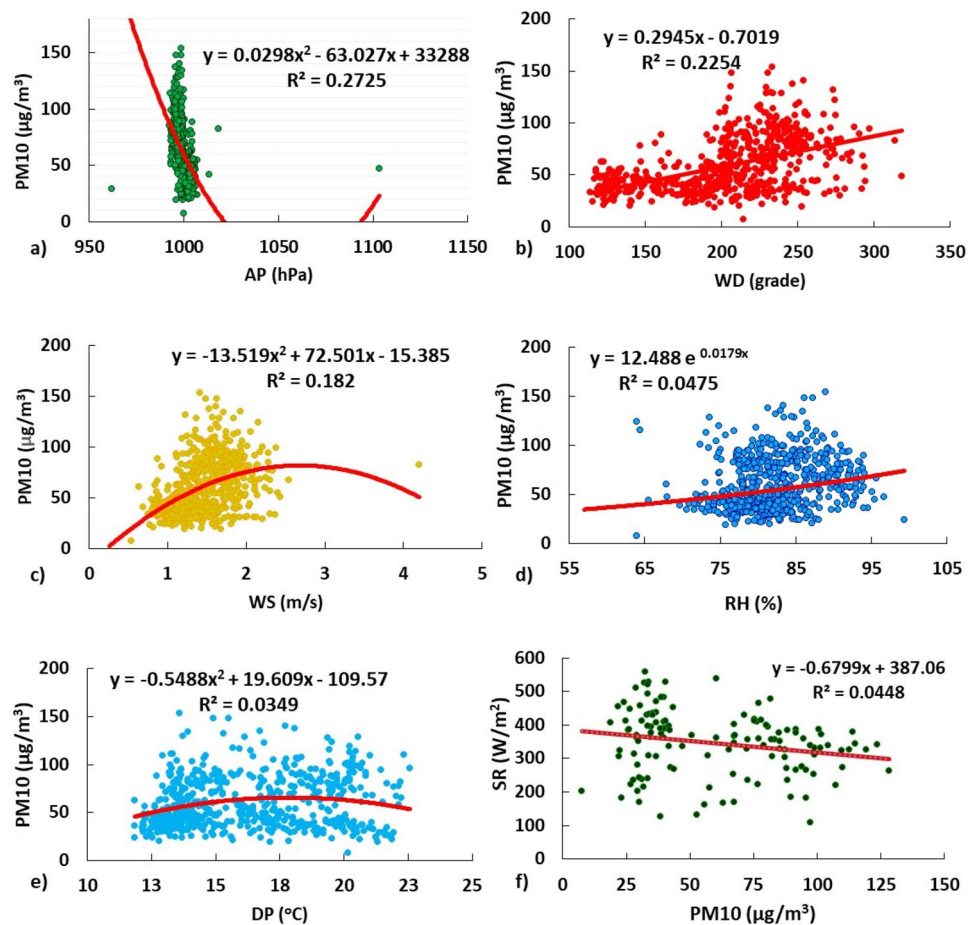
the range of velocities was lower for SS (1.15 – 1.34 m/s) and higher for SJM (1.49–1.86 m/s). Regarding the WS-PM<sub>10</sub> relationship, some authors calculated negative correlations<sup>13,14</sup>. On the contrary, positive and direct correlations were calculated ( $r_{WS-PM_{10}} - \text{annual}(0.40526)$ ), especially for summer ( $r_{WS-PM_{10}}(0.55296)$ ) and autumn ( $r_{WS-PM_{10}}(0.56274)$ ). These values are higher than other studies reported in Lima<sup>34</sup>, indicating influence on PM<sub>10</sub> dispersion, resuspension and transport, including its decrease with the simultaneous diminution of wind<sup>50</sup>. The non-parametric regression produced a coefficient of determination ( $R^2 = 0.182$ ), close to other studies on the same variable<sup>13</sup> (Fig. 6c). On the other hand, with respect to annual WD in southern Lima, the correlations were also significant  $r_{WD-PM_{10}}(0.48192)$  (Fig. 5b), in the order:  $r_{spring}(0.72592) > r_{summer}(0.55989) > r_{winter}(0.5149) > r_{autumn}(0.08837)$ .



**Figure 5.** Daily variations of meteorological variables and concentrations of  $PM_{10}$ , for South Lima in 2019. (a) WS, (b) WD, (c) AP, (d) DP, (e) SR and (f) RH.

Regarding the atmospheric pressure presented in Fig. 5c, shows higher AP values for lower  $PM_{10}$  values, confirmed by their negative correlation in the order: summer ( $r_{AP-PM_{10}} = -0.74682$ ) > autumn ( $r_{AP-PM_{10}} = -0.72994$ ) > spring ( $r_{AP-PM_{10}} = -0.54587$ ) > annual ( $r_{AP-PM_{10}} = -0.39443$ ) (see Table 2). In that sense, Li points out that low atmospheric pressures associated with downward mass fluxes restrict the upward movement of PM by accumulating them in the air column<sup>14</sup>. The AP- $PM_{10}$  statistical adjustment produced a weak significant regression ( $R^2 = 0.2725$ ,  $p < 0.05$ ), being a stronger relationship calculated with respect to the other variables.

On the other hand, temperature, dew point and relative humidity were also evaluated in this study. Contrary to Govindasamy's study, no significant correlations were generated for T- $PM_{10}$ , but significant



**Figure 6.** Regressions of a single meteorological factor ( $p < 0.05$ ) and  $PM_{10}$  in South Lima in 2019: (a) AP- $PM_{10}$  Annual, (b) WD- $PM_{10}$  Annual, (c) WS- $PM_{10}$  Annual, (d) RH- $PM_{10}$  Annual, (e) DP- $PM_{10}$  Annual, (f)  $PM_{10}$ -SR (summer).

correlations were generated for DP- $PM_{10}$ <sup>49</sup>. These were weak and direct in winter ( $r_{DP-PM_{10}} = 0.2957$ ) and spring ( $r_{DP-PM_{10}} = 0.184$ ), and the regressions were also weak ( $R^2 = 0.0349$ ) (Fig. 6e). According to Szep, if  $T > DP$ , as occurs in southern Lima, the pollutant concentration and atmospheric stability decrease causing dilution of the pollutant and its partial wet precipitation<sup>40</sup>. The SS zone presents a greater  $T - DP$  (SS :  $3.65^\circ\text{C}$ ) difference than SJM (SJM :  $2.59^\circ\text{C}$ ) which would favor a greater dilution of  $PM_{10}$ .

On the other hand, it is evident that the high RH in Lima in winter is associated with haze and rainfall events, generating drifting particulate matter ( $PM_{10}$ ), that is removed from the air column by wet precipitation<sup>40,51</sup>. The warm periods (summer) do not usually present rainfall and RH decreases as a characteristic of desert geography, generating a significant T-RH correlation ( $r_{T-RH} - \text{summer} = 0.67751$  and  $r_{T-RH} - \text{summer} = 0.80644$ ). A more stable atmosphere with progressive decrease in the altitude of the TI base in January-May (warm periods) produced greater evaporation and accumulation of  $PM_{10}$  especially in SJM. The RH- $PM_{10}$  correlation presented the following order: autumn ( $r_{RH-PM_{10}} = 0.22361$ ) > spring ( $r_{RH-PM_{10}} = 0.39192$ ) > winter ( $r_{RH-PM_{10}} = 0.35943$ ), which generated a weak regression ( $R^2 = 0.0475$ , Fig. 6d) as found in other international studies ( $r_{RH-PM_{10}} = 0.382$  to  $0.467$ )<sup>15</sup>.

In addition, about the solar radiation presented in Fig. 4g shows a marked variability in the trends between SR and  $PM_{10}$ , with a weak inverse correlation in summer ( $r = -0.21155$ ). This led to calculate particular correlations in each zone according to the seasons of the year, among them: summer ( $r_{SS} = 0.15194$ ;  $r_{SJM} = 0.009844$ ), autumn ( $r_{SS} = 0.03957$ ;  $r_{SJM} = 0.124166$ ), winter ( $r_{SS} = 0.28932$ ,  $r_{SJM} = 0.5552$ ) and spring ( $r_{SS} = 0.33422$ ;  $r_{SJM} = 0.44675$ ). The results were congruent with those reported in other investigations; for example, Vardoulakis and Kassomenos, also reported weak correlations in European cities during warm ( $r = 0.02$  to  $0.06$ ) and cool ( $r = 0.11$  to  $0.38$ ) months<sup>52</sup>. On the other hand, direct combinations of high  $PM_{10}$ -SR or low  $PM_{10}$ -SR values would be infrequent and suggest that high  $PM_{10}$  concentrations could decrease RH intensity<sup>53</sup>. Indeed, the study showed reductions in RH in this coastal area of the southern solstice circle<sup>54</sup>, being between 11.5% (summer) and 25.3% (winter) for elevated seasonal averages of PM in SJM ( $73.14$  to  $87.21 \mu\text{g}/\text{m}^3$ ) compared to SS, and mineral dust aerosols scatter and absorb some of the RH reaching land<sup>53</sup>.

	Summer			Autumn			Winter			Spring			Annual		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
EV	2.1	2	1	2.6	1.6	1	2.3	1.5	0.9	2.6	2	0.7	2.7	1.5	0.8
PV	35	32	16	44	27	16	39	25	15	43	33	12	44	26	14
CV	35	67	83	44	71	88	39	64	80	43	76	88	44	70	84
RH	0.43	-0.4	-0.2	-0.5	0.3	-0.02	0.57	-0.2	0.25	0.37	-0.45	0.14	-0.3	0.58	0
T	-0.21	0.7	0.1	0.6	-0.1	0.15	-0.07	0.76	-0.31	0.15	0.68	-0.14	0.58	-0.22	-0.1
DP	0.13	0.6	0.1	0.5	0.1	0.2	0.48	0.49	-0.02	0.39	0.5	-0.08	0.58	-0.05	-0.13
AP	-0.6	-0.3	-0.1	0	-0.7	0.1	-0.2	0.27	0.9	-0.5	0.14	0.5	-0.3	-0.48	0.01
WS	0.42	0.2	-0.7	0.3	0.6	-0.3	0.41	0.18	0.16	0.4	0.12	0.83	0.29	0.26	0.87
WD	0.46	0	0.7	-0.2	0.3	0.92	0.48	-0.19	-0.1	0.52	-0.21	-0.16	0.25	0.55	-0.45

**Table 3.** Principal Components of meteorological variables. EV: Eigenvalue (%); PV: Percentage of variance (%) and CV: Cumulative variance.

**Multivariate relationships using the PCA.** Table 3 shows the principal component analysis. Eigenvalues were produced for three principal components (PC) that explained between 80% and 88% of the total variance of PM<sub>10</sub> concentrations. As the PC factor is the square of the factor loading, it has been interpreted as the equivalent of the coefficient of determination. The PM<sub>10</sub> variances in summer (83%), autumn (88%), winter (80%) and spring (88%) showed moderate loadings of the variables with only one factor<sup>55</sup>.

In summer, factors PC1 (RH-AP) and PC2 (T-DP) explained 67% of the variance, with descending wind flows and low pressure levels favoring increased humidity and atmospheric stability and decreasing the altitude of the TI base and thermal gradients (1.1 to 0.9°C/100 m), increasing PM<sub>10</sub><sup>56</sup>. In autumn, factors PC1 (RH-T-DP) and PC2 (AP-WS) explained 71% of the variance, witnessing in May atmospheric stability that prevented vertical development of the mixing layer and maintained high PM<sub>10</sub> levels, but in June coastal winds, humidity, thermal gradient (2.5°C/100 m) and TI base altitude (756.6 m) intensified<sup>56</sup>, driving aerosol dispersion. In winter, factors PC1 (RH-WS-WD) and PC2 (T-DP) explained 64% of the variance, with higher thermal gradients (2.6°C – 3.4°C/100 m), wind intensity, humidity and TI base altitude (> 750 m) favoring higher PM<sub>10</sub> dispersion<sup>56</sup>. In spring, factors PC1 (AP-WD) and PC2 (RH-T-DP), explained 76% of the variance, producing greater atmospheric instability without significant humidity inputs and temperature increases that warmed the surface and favored the dispersion of the pollutant<sup>56</sup>. Figure 7 shows the 2 main factors (PC1 and PC2) that concentrated the highest percentages of the PCA in the seasons of the year for 2019.

**Multiple regression model.** The meteorological and PM<sub>10</sub> data observed in 2019 produced the following multiple regression model with significant p-value (p < 0.05):

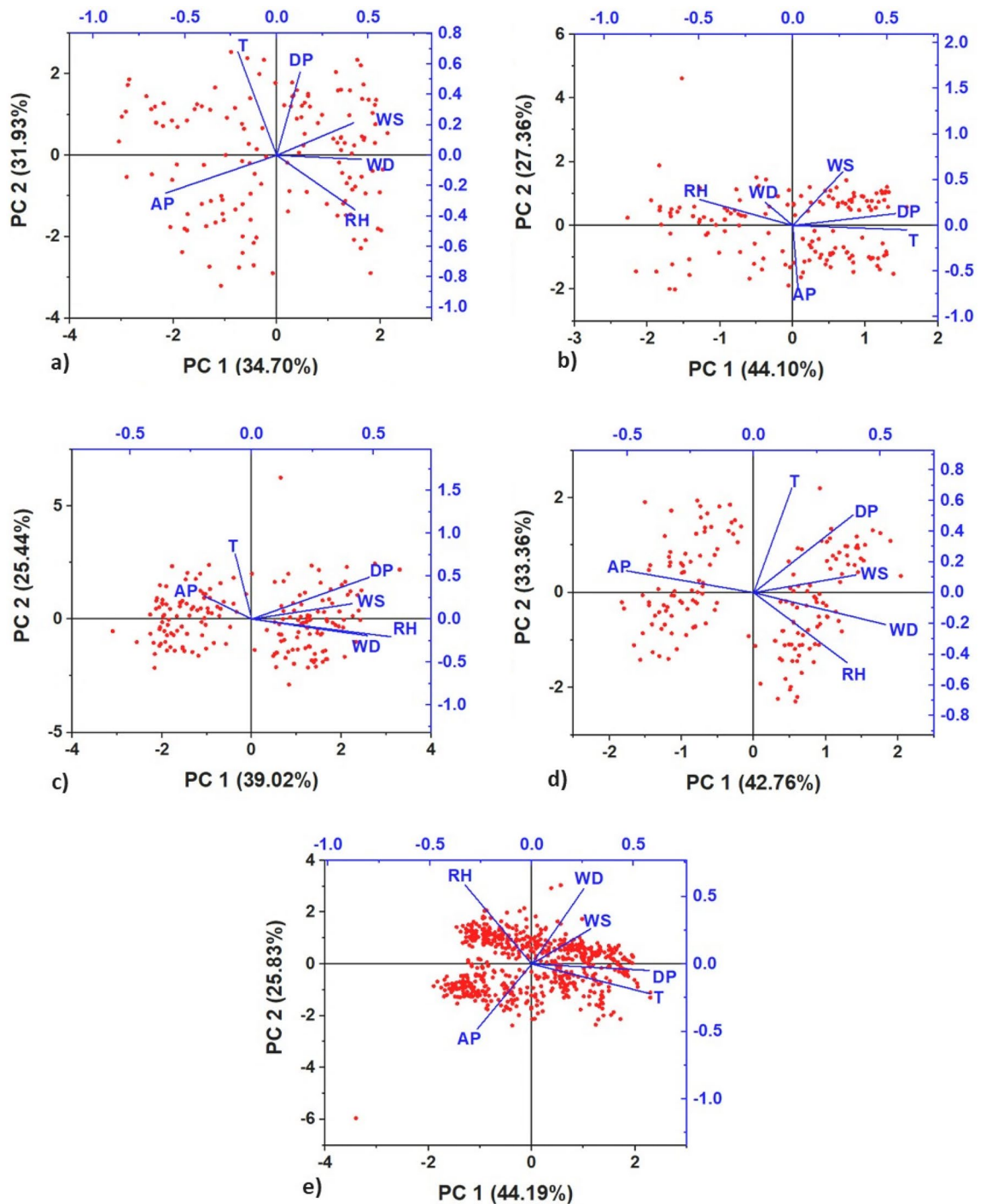
$$PM_{10} = 612.9611 + 2.90988RH + 14.68703T - 16.8064DP - 0.88883AP + 22.90704WS + 0.251WD \quad (13)$$

The relative error for the intercept was equal to 217.335 and for the coefficients of the variables ranged from 0.16989 (AP) to 6.5235 (DP). The determination factor reflected a weak fit ( $R^2 = 0.3802$ ), evidencing the limitations of the method. This result was consistent with the model described by Lin<sup>38</sup>, for the T-WS-PM<sub>10</sub> combination ( $R^2 = 0.394$ ). In contrast, Ceylan did not obtain a significant model in similar tests<sup>39</sup>. Likewise, Kamarul related RH-T with a MODIS-AOD550 satellite factor to improve the regression ( $R^2 = 0.66$ ), but suggested optimizing it and including WS-WD<sup>16</sup>.

**Three-dimensional models.** The multiple variables integrally influence the dilution and diffusion of atmospheric pollutant<sup>13</sup>. Under this assumption, statistical fitting of 3D surfaces was performed. The results yielded strong and significant regressions ( $R^2 > 0.75$ ; p < 0.05), in the order: RH – AP – PM<sub>10</sub> ( $R^2 = 0.94685$ , Fig. 8a) > T – DP ( $R^2 = 0.87646$ , Fig. 8b) > (AP – WS – PM<sub>10</sub>,  $R^2 = 0.85064$ , Fig. 8c) > (AP – WD – PM<sub>10</sub>,  $R^2 = 0.77984$ , Fig. 8d). These three-dimensional models (see Table 4) correspond with the results obtained in the PCA and explain that relative humidity and atmospheric pressure largely affect the PM<sub>10</sub> concentration, as well as temperatures and wind action. Compared with the curve fitted under the influence of a single factor and MLR, the fitting performance of the functional relationship is higher and confirms that different meteorological factors have different effects on PM<sub>10</sub> concentration.

**Comparing model predictions to monitoring data 2019 and model applied to 2020.** Figure 9 shows the data fit between the observed PM<sub>10</sub> and the calculated (modeled) PM<sub>10</sub> for the year 2019. While, Fig. 10, represents the data fit in the 2020 assessment. The calibration of the models evaluated by the correlation coefficient indicates that the MLR performed better than the other models developed (r = 0.6166) between the outputs (modeled PM<sub>10</sub>) and the observed data (Fig. 9a). Also, the three-dimensional function that combined the LogAP-LogWD-LogPM<sub>10</sub> (Fig. 9b) presented a moderate correlation (r = 0.5753) unlike the other two 3D models. However, Table 5 shows that the RMSE for the MLR was higher (RMSE= 12.9226) relative to the others, but comparable to the modeling errors of another study (RMSE= 10.64 – 26.08, T – AP – RH – WS)<sup>45</sup>. The 3D models had smaller errors (RMSE = 0.0989 a 0.2776) because the algorithm for 3D models by regres-



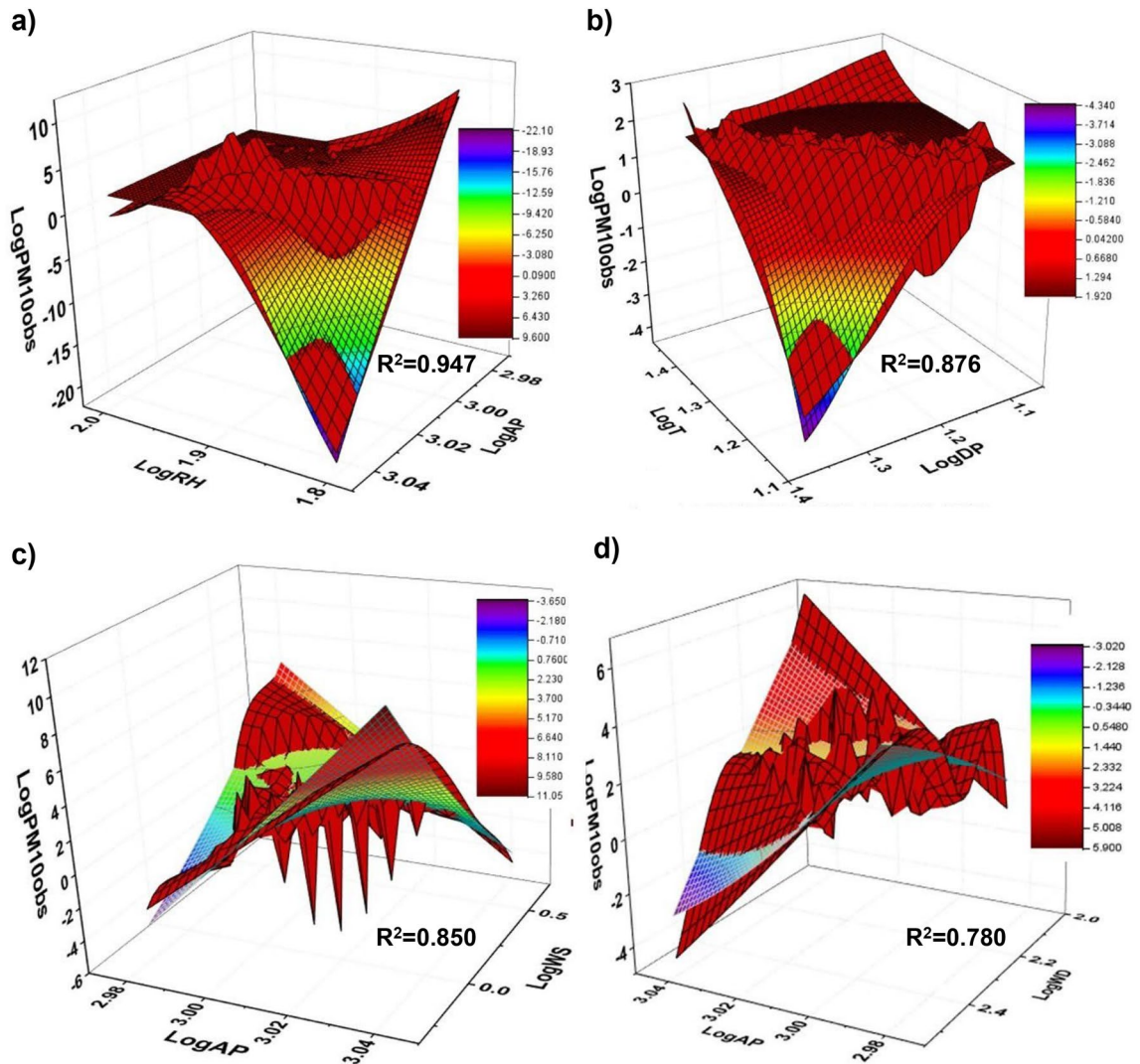


**Figure 7.** Principal components analysis: (a) summer, (b) autumn, (c) winter, (d) spring, (e) annual.

sion fitting generates more complex interactions between input and output data. The NSE criterion for MLR (NSE=0.3804) was closer to unity, and was also comparable to Nguyen's fitting errors<sup>45</sup> (between 0.26 and 0.53).

The evaluation of the models applied to the 2020 data showed a decrease in correlations, with the exception of the combined 3D model of RH-AP-PM<sub>10</sub> and MLR which showed slightly higher correlations with ( $r = 0.4435$ ) and ( $r = 0.3239$ ) respectively. The RMSE for the MLR doubled relative to its calibration (RMSE = 23.9983) and the values of the NSE criterion were all negative, but the NSE of the MLR and the 3D combined AP-WS models were closer to unity error, with values of  $-2.7214$  and  $-1.5304$ , respectively. Consequently, two relevant aspects were highlighted:

- Conditions in 2019 were characterized by intense anthropogenic activity. While in 2020, the cessation of activities at the beginning of the pandemic showed changes during the months of blocking and subsequent reactivation (May). These changes were reflected in the wind patterns, especially for SS, which presented a



**Figure 8.** Functional relationships for meteorological bi-variable combinations with PM<sub>10</sub> concentration, expressed as: (a) logRH-logAP, (b) logT-logDP, (c) logAP-logWS and (d) logAP-logWD.

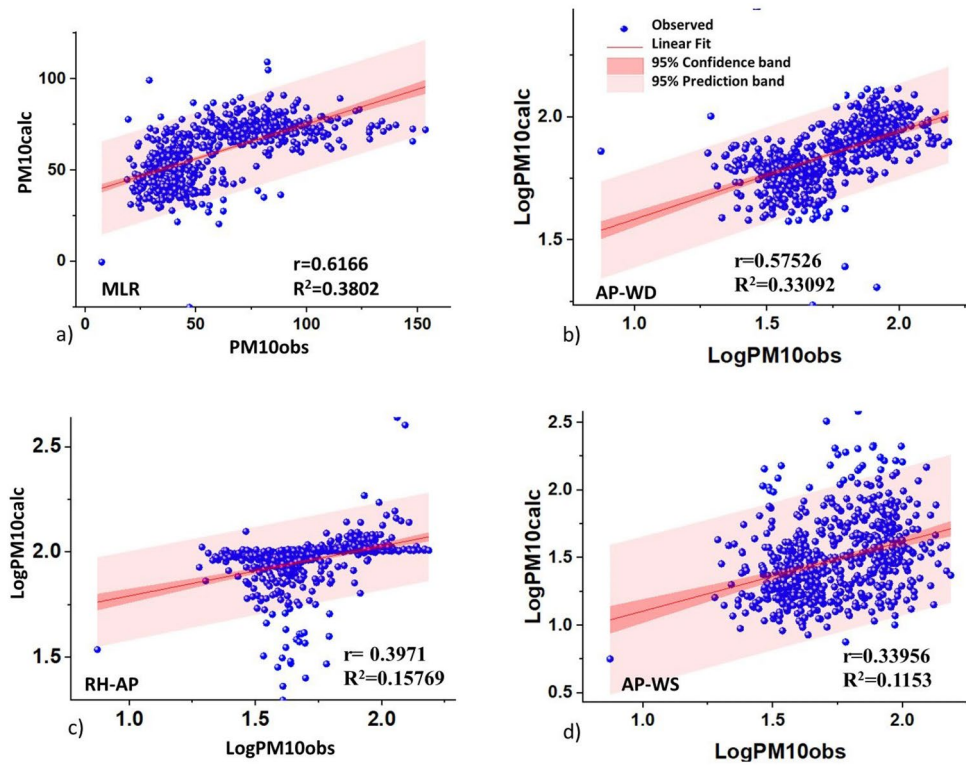
Model	Notation
M1	$z = -0.29 - 311.14e^{-e^{-(x-1.85)/(-0.05)}} + 3.17e^{-e^{-(y-3.15)/(-0.14)}} + 428.35e^{-e^{-(x-1.85)/(-0.05)}} - e^{-(y-3.15)/(-0.14)}$
M2	$z = 1.96 - 1.03(0.014/((4(x - 1.06)^2 + 0.05)(4(y - 1.38)^2 + 0.06))) + 8.84 \ln 2 e^{(-4 \ln 2 / 0.05)(x - 1.06)^2 - (4 \ln 2 / 0.07)(y - 1.38)^2}$
M3	$z = 1204.85 - 913.03x + 1391.56y + 170.58x^2 + 3.62y^2 - 463.20xy$
M4	$z = 3139.14 - 3175.62x + 1469.88y + 706.62x^2 - 5.50y^2 - 481.38xy$

**Table 4.** Three-dimensional models. M1: Extreme Cum; M2: Voigt2DMod; M3: Poly2D and M4: Poly2D.

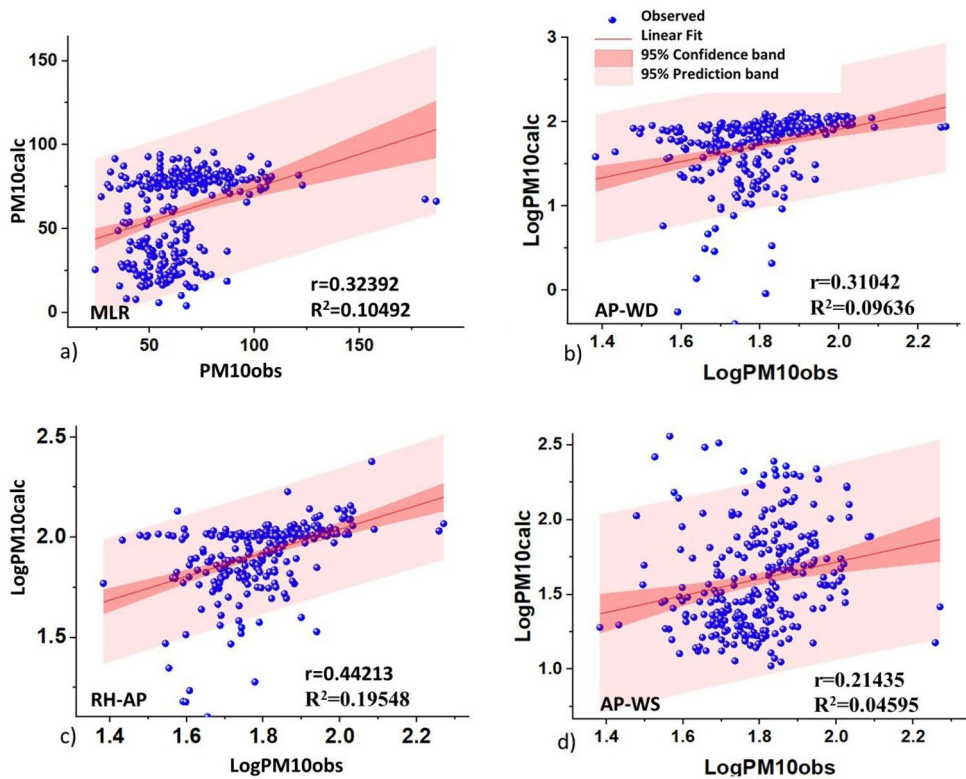
dominant direction towards the south in 2019. While in 2020 it was from east to south east, associated with the increase in PM<sub>10</sub>.

- The results should be considered significant for predicting PM<sub>10</sub> concentration. However, it is believed that the inclusion of new predictor variables related to TI base altitude, aerosol re-suspension, vehicular traffic and discriminations of anthropogenic and geogenic sources<sup>57</sup> would help to improve the model to compare with others in order to minimize human health risks in times of pandemic.

**Comparison of research on atmospheric quality in the city of Lima, Peru.** The results of research conducted by different authors on air quality related to PM<sub>10</sub> in the city of Lima were compared with the present study.



**Figure 9.** Calibrations 2019: (a) multiple linear regression, (b) Log AP- Log WS-Log PM<sub>10</sub> combinations, (c) Log RH-Log AP-Log PM<sub>10</sub>, (d) Log AP-Log WS-Log PM<sub>10</sub>.



**Figure 10.** Evaluation 2020: (a) Multiple Linear Regression, (b) Log AP- Log WS-Log PM<sub>10</sub> combinations, (c) Log RH-Log AP-Log PM<sub>10</sub>, (d) Log AP-Log WS-Log PM<sub>10</sub>.

Metrics								
Model	Pearson's r		R <sup>2</sup>		RMSE		NSE	
	CAL	EVA	CAL	EVA	CAL	EVA	CAL	EVA
MLR	0.6166	0.3239	0.3802	0.1049	12.9226	23.9983	0.3804	- 2.7214
Model 1	0.5753	0.3019	0.3309	0.0911	0.0989	0.3773	0.0544	- 8.0018
Model 2	0.3971	0.4435	0.1577	0.1967	0.1065	0.1545	- 1.2841	- 1.5304
Model 3	0.3396	0.2144	0.1153	0.046	0.2776	0.3305	- 2.9502	- 9.0345

**Table 5.** Statistical comparison results between the models studied for PM<sub>10</sub>. Model 1: Model 3D (LogAP-LogWD-LogPM<sub>10</sub>); Model 3D (LogRH-LogAP-LogPM<sub>10</sub>); Model 3D (LogAP-LogWS-LogPM<sub>10</sub>); CAL: Calibration and EVA: Evaluation.

- This research provides an easy and practical method with effective and reliable results through the development of statistical prediction models for PM<sub>10</sub>, based on multiple linear regression, use of three-dimensional logarithms and principal component analysis, under the influence of meteorological variables in the warm and cool season in southern Lima. This technique allows testing the applicability of the models and reveals the spatial distribution dynamics of PM<sub>10</sub>, strengthening decision making in environmental management related to the protection of human health through the prevention and control of PM<sub>10</sub> air pollution in the context of constant urban growth.
- Silva et al.<sup>58</sup> evaluated the PM<sub>10</sub> pollutant in the city of Lima over a 6-year period (2010–2015), showing that the highest PM<sub>10</sub> concentrations were observed in the eastern part of the city, mainly in the summer (December to March). In addition, the authors identified large open spaces, vehicular traffic and the commercialization of rubble, bricks and cement as the main sources of particulate matter. These results are similar to those reported in this research; however, the authors conducted the research in a period before the COVID-19 pandemic, which reflects a stable situation in environmental conditions.
- Reátegui-Romero et al.<sup>59</sup> conducted a study on PM<sub>10</sub> and PM<sub>2.5</sub> pollutants during 2 months (February and July 2016), showing that the highest PM<sub>10</sub> concentrations were observed in the northern area of Lima, and relative humidity is inversely proportional to PM<sub>10</sub> concentrations, with higher peaks observed in the summer month (February). The authors' results coincide with the findings of this research for such months, which are explained through the seasonal behavior pattern of South Lima and in the 3D model that demonstrates the influence of association between RH and AP on PM<sub>10</sub> (Fig. 8a).
- Sanchez et al.<sup>12</sup> used the WRF-Chem to predict PM<sub>10</sub> concentrations in Lima during April 2016, showing that there is a higher PM<sub>10</sub> concentration in areas with greater impact of vehicular traffic, reaching values of 476.8 µg/m<sup>3</sup> for the *Santa Anita* station. The authors related temperature, relative humidity and wind speed, in addition to the incorporation of topographic and meteorological data that increased the accuracy in terms of normalized mean bias for PM<sub>10</sub> based on the emissions inventory. In contrast, this research was based on the modeling of air quality through the exclusive relation of meteorological variables with PM<sub>10</sub>, showing the limitations of the applied models that could be enhanced with the inclusion of geomorphological factors, among others.
- Cordova et al.<sup>34</sup> evaluated the PM<sub>10</sub> pollutant in Metropolitan Lima during 2017 and 2018, mentioning that the main sources were the vehicle fleet, the industrial park and overpopulation, reaching maximum values (974 µg/m<sup>3</sup>) at the *Huachipa* station for the summer months (December–March). Artificial neural networks, specifically, the Long Short-Term Memory model under two validation schemes were used to predict PM<sub>10</sub> concentrations. The results showed good prediction performance for both low concentrations and critical episodes. The model presented a potential application for South Lima and could be compared with the simple methodology (MLR, 3D and PCA) applied in this research. However, an analysis of the RMSE errors calculated in both applications resulted in discrepant values (statistical methods: 0.0989 to 23.9983; ANN: 10.573 to 64.297) and very close Spearman correlations (MLR: 0.6166; 3D RH-AP Model: 0.5753; ANN: 0.517–0.756).

### Limitations

This study has some limitations. There are gaps, about 67% are valid data (2019–2020). It is also limited to the application of statistical functions such as MRL, PCA and logarithmic functions using the factors, coefficient of determination, correlation coefficient, RMSE and NSE. The model is proposed with data from 2019 and extrapolated to the following year due to the limited availability of data in 2020. The application of three-dimensional models is limited by their low R<sup>2</sup> (for the year 2020). The number of data represents a relatively short period (two years). A more extended period of hourly data may have allowed a more rigorous statistical analysis and more conclusive results.

### Conclusions

A statistical modeling approach has been applied to predict PM<sub>10</sub> concentration in two locations in South Lima (SJM and SS) both before (2019) and during the COVID-19 pandemic (2020), as a function of meteorological variables. The PCA evidenced the seasonal influence explained in the various combinations of meteorological variables on the distribution of PM<sub>10</sub>. The SJM district presented moderate to poor PM<sub>10</sub> quality levels versus most acceptable values in SS. Calibration of the statistical models in 2019 demonstrated a better (significant) fit



for the multiple linear regression model than the 3D modeling, while evaluation of the models in 2020 generated lower determination factors. Thus, this research strengthens the application of statistical models in predicting the spatial distribution of PM<sub>10</sub>, providing scientific support in decision making related to public health protection. As future work, we consider developing new models under the machine learning approach through the application of geostatistical models to compare the accuracy in the prediction of air quality for PM<sub>10</sub>. Likewise, address the analysis of air pollution before, during and after the pandemic using diagnostic measures for the class of nonparametric regression models with symmetric random errors, which includes all continuous and symmetric distributions<sup>60</sup>.

## Data availability

The collection and statistical processing of the data was carried out under the authorization of *Dirección General de Salud Ambiental*, the governmental entity responsible for the sanitary surveillance of air quality and meteorology. The datasets are available in the repository, [http://www.digesa.minsa.gob.pe/DEPA/aire\\_lc/lima\\_callao.asp](http://www.digesa.minsa.gob.pe/DEPA/aire_lc/lima_callao.asp).

Received: 24 May 2022; Accepted: 20 September 2022

Published online: 06 October 2022

## References

- Liu, N., Liu, X., Jayaratne, R. & Morawska, L. A study on extending the use of air quality monitor data via deep learning techniques. *J. Clean. Prod.* **274**, 122956. <https://doi.org/10.1016/j.jclepro.2020.122956> (2020).
- Amann, M. *et al.* Managing future air quality in megacities: A case study for Delhi. *Atmos. Environ.* **161**, 99–111. <https://doi.org/10.1016/j.atmosenv.2017.04.041> (2017).
- Wang, Z., Delp, W. W. & Singer, B. C. Performance of low-cost indoor air quality monitors for pm2.5 and pm10 from residential sources. *Build. Environ.* **171**, 106654. <https://doi.org/10.1016/j.buildenv.2020.106654> (2020).
- Maesano, C. *et al.* Impacts on human mortality due to reductions in pm10 concentrations through different traffic scenarios in Paris, France. *Sci. Total Environ.* **698**, 134257. <https://doi.org/10.1016/j.scitotenv.2019.134257> (2020).
- Gryparis, A., Dimakopoulou, K., Pedeli, X. & Katsouyanni, K. Spatio-temporal semiparametric models for no2 and pm10 concentration levels in Athens, Greece. *Sci. Total Environ.* **479**, 21–30. <https://doi.org/10.1016/j.scitotenv.2014.01.075> (2014).
- Sturm, R. Modeling the deposition of bioaerosols with variable size and shape in the human respiratory tract: A review. *J. Adv. Res.* **3**, 295–304. <https://doi.org/10.1016/j.jare.2011.08.003> (2012).
- Dastoorpoor, M. *et al.* Exposure to particulate matter and carbon monoxide and cause-specific cardiovascular-respiratory disease mortality in Ahvaz. *Toxicol. Res.* **40**, 1362–1372. <https://doi.org/10.1080/15569543.2020.1716256> (2021).
- Kim, H.-B., Shim, J.-Y., Park, B. & Lee, Y.-J. Long-term exposure to air pollutants and cancer mortality: A meta-analysis of cohort studies. *Int. J. Environ. Res. Public Health* **15**, 2608. <https://doi.org/10.3390/ijerph15112608> (2018).
- Song, C. *et al.* Air pollution in China: Status and spatiotemporal variations. *Environ. Pollut.* **227**, 334–347. <https://doi.org/10.1016/j.envpol.2017.04.075> (2017).
- WHO. Who air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide: Global update 2005: Summary of risk assessment. type Tech. Rep., institution World Health Organization (2006).
- Falocchi, M. *et al.* A dataset of tracer concentrations and meteorological observations from the Bolzano tracer experiment (btex) to characterize pollutant dispersion processes in an alpine valley. *Earth Syst. Sci. Data* **12**, 277–291. <https://doi.org/10.5194/essd-12-277-2020> (2020).
- Sánchez-Ccoyllo, O. R. *et al.* Modeling study of the particulate matter in lima with the wrf-chem model: Case study of April 2016. *Int. J. Appl. Eng. Res. IJAER* **13**, 10129. <https://doi.org/10.37622/IJAER/13.11.2018.10129-10141> (2018).
- Qi, X., Mei, G., Cuomo, S., Liu, C. & Xu, N. Data analysis and mining of the correlations between meteorological conditions and air quality: A case study in Beijing. *Internet Things* **14**, 100127. <https://doi.org/10.1016/j.iot.2019.100127> (2021).
- Li, X., Ma, Y., Wang, Y., Liu, N. & Hong, Y. Temporal and spatial analyses of particulate matter (pm10 and pm2.5) and its relationship with meteorological parameters over an urban city in Northeast China. *Atmos. Res.* **198**, 185–193. <https://doi.org/10.1016/j.atmosres.2017.08.023> (2017).
- Zheleva, I., Tsvetanova, I. & Filipova, M. Statistical study of the influence of the atmospheric characteristics upon the particulate matter (pm10) air pollutant in the city of Silistra, Bulgaria. In *AIP Conference Proceedings*, 120014, <https://doi.org/10.1063/1.5130874> (AIP Publishing LLC, 2019).
- Zaman, N. A. F. K., Kanniah, K. D. & Kaskaoutis, D. G. Estimating particulate matter using satellite based aerosol optical depth and meteorological variables in Malaysia. *Atmos. Res.* **193**, 142–162. <https://doi.org/10.1016/j.atmosres.2017.04.019> (2017).
- Safar, Z., Labib, M. W. & Gertler, A. W. Development and validation of a lead emission inventory for the greater Cairo area. *J. Adv. Res.* **5**, 551–562. <https://doi.org/10.1016/j.jare.2013.07.003> (2014).
- Velásquez, R. A., Ramos, Y. R. & Noel, J. Citizen science approach for spatiotemporal modelling of air pollution quality and traffic in Lima, Peru. In *2019 IEEE Sciences and Humanities International Research Conference (SHIRCON)*, 1–4, <https://doi.org/10.1109/SHIRCON48091.2019.9024879> (IEEE, 2019).
- Martelletti, L. & Martelletti, P. Air pollution and the novel covid-19 disease: A putative disease risk factor. *SN Compr. Clin. Med.* **2**, 383–387. <https://doi.org/10.1007/s42399-020-00274-4> (2020).
- Zangari, S., Hill, D. T., Charette, A. T. & Mirowsky, J. E. Air quality changes in New York city during the covid-19 pandemic. *Sci. Total Environ.* **742**, 140496. <https://doi.org/10.1016/j.scitotenv.2020.140496> (2020).
- Rojas, J. P. *et al.* Effects of covid-19 pandemic control measures on air pollution in lima metropolitan area, Peru in South America. *Air Qual. Atmos. Health* **14**, 925–933. <https://doi.org/10.1007/s11869-021-00990-3> (2021).
- MINAM. Informe nacional sobre el estado del ambiente 2014–2019. Tech. Rep., Ministerio del Ambiente (2021).
- ATU. Indicadores-atu reporte de viajes diarios de lima y callao al año 2019. Tech. Rep., Autoridad de Transporte Urbano (2021).
- MINAM. Aprueban estándares de calidad ambiental (eca) para aire y establecen disposiciones complementarias decreto supremo no 003-2017-minam. Tech. Rep., Ministerio del Ambiente (2017).
- MINAM. Resolución ministerial 181-2016-minam. Tech. Rep., Ministerio del Ambiente (2016).
- INEI. Perú: Estimaciones y proyecciones de población por departamento, provincia y distrito, 2018–2020. Tech. Rep., Instituto Nacional de Estadística e informática (2020).
- Adams, M. D. & Kanaroglou, P. S. Mapping real-time air pollution health risk for environmental management: Combining mobile and stationary air pollution monitoring with neural network models. *J. Environ. Manag.* **168**, 133–141. <https://doi.org/10.1016/j.jenvman.2015.12.012> (2016).
- Croituru, C. & Nastase, I. A state of the art regarding urban air quality prediction models. *E3S Web Conf.* **32**, 01010. <https://doi.org/10.1016/j.jenvman.2015.12.012> (2018).

29. Li, X., Peng, L., Hu, Y., Shao, J. & Chi, T. Deep learning architecture for air quality predictions. *Environ. Sci. Pollut. Res.* **23**, 22408–22417. <https://doi.org/10.1007/s11356-016-7812-9> (2016).
30. Salini Calderón, G. & Pérez Jara, P. Estudio de series temporales de contaminación ambiental mediante técnicas de redes neuronales artificiales. *Ingeniare. Revista chilena de ingeniería* **14**, 284–290 (2006).
31. Guzmán, A. A. E. *et al.* Artificial neural network modeling of PM10 and PM2.5 in a tropical climate region: San Francisco de Campeche, Mexico. *Química Nova* **40**, 1025–1034. <https://doi.org/10.21577/0100-4042.20170115> (2017).
32. Delgado, A., Montellanos, P. & Llave, J. Air quality level assessment in Lima city using the grey clustering method. In *2018 IEEE International Conference on Automation/XXIII Congress of the Chilean Association of Automatic Control (ICA-ACCA)*, 1–4. <https://doi.org/10.1109/ICA-ACCA.2018.8609699> (IEEE, 2018).
33. Carmen, G., Rita, C., Patricia, R. & Lorgio, V. Dependencia del material particulado en aire respecto a parámetros meteorológicos en la localidad de zárate (sjl). Tech. Rep. number13, Tecsup (2019).
34. Cordova, C. H. *et al.* Air quality assessment and pollution forecasting using artificial neural networks in metropolitan Lima-Peru. *Sci. Rep.* **11**, 1–19. <https://doi.org/10.1038/s41598-021-03650-9> (2021).
35. Wang, P. *et al.* Responses of pm2.5 and o3 concentrations to changes of meteorology and emissions in China. *Sci. Total Environ.* **662**, 297–306. <https://doi.org/10.1016/j.scitotenv.2019.01.227> (2019).
36. He, J. *et al.* Influences of meteorological conditions on interannual variations of particulate matter pollution during winter in the Beijing–Tianjin–Hebei area. *J. Meteorol. Res.* **31**, 1062–1069. <https://doi.org/10.1007/s13351-017-7039-9> (2017).
37. Kermani, M. *et al.* Characterization, possible sources and health risk assessment of pm2.5-bound heavy metals in the most industrial city of Iran. *J. Environ. Health Sci. Eng.* **19**, 151–163. <https://doi.org/10.1007/s40201-020-00589-3> (2021).
38. Lin, C.-Y., Chiang, M.-L. & Lin, C.-Y. Empirical model for evaluating pm10 concentration caused by river dust episodes. *Int. J. Environ. Res. Public Health* **13**, 553. <https://doi.org/10.3390/ijerph13060553> (2016).
39. Ceylan, Z. & Bulkan, S. Forecasting pm10 levels using ann and mlr: A case study for Sakarya city. *Glob. Nest J.* **20**, 281–90 (2018).
40. Szep, R. *et al.* Dew point-indirect particulate matter pollution indicator in the Ciuc basin-Harghita, Romania. *Rev. Chim.* **67**, 1914–1921. <https://doi.org/10.1016/j.jare.2011.08.003> (2016).
41. Multisectorial, C. Avance: Plan de acción para el mejoramiento de la calidad del aire de lima-callao, diagnóstico de la gestión de la calidad ambiental del aire de lima y callao. Tech. Rep., Comisión Multisectorial para la gestión de la iniciativa del aire limpio para Lima y Callao (2019).
42. Valdivia, S. A. P. Análisis temporal y espacial de la calidad del aire determinado por material particulado pm10 y pm2.5 en lima metropolitana. *Anales Científicos* **77**, 273–283. <https://doi.org/10.21704/ac.v77i2.699> (2016).
43. DIGESA. Protocolo de monitoreo de la calidad del aire y gestión de los datos. Tech. Rep., Dirección General de Salud Ambiental (2005).
44. Kherif, F. & Latypova, A. Chapter 12—Principal component analysis. In *Machine Learning* (eds Mechelli, A. & Vieira, S.) 209–225 (Academic Press, 2020). <https://doi.org/10.1016/B978-0-12-815739-8.00012-2>.
45. Nguyen, D. A., Duong, S. H., Tran, P. A., Cao, H. H. & Ho, B. Q. Combination of data-driven models and interpolation technique to develop of pm10 map for Hanoi, Vietnam. *Sci. Rep.* **10**, 1–14. <https://doi.org/10.1038/s41598-020-75547-y> (2020).
46. Azad, A. *et al.* Novel approaches for air temperature prediction: A comparison of four hybrid evolutionary fuzzy models. *Meteorol. Appl.* **27**, e1817. <https://doi.org/10.1002/met.1817> (2020).
47. SENAMHI. Atlas de energía solar del Perú. proyecto per/98/g31 “electrificación rural a base de energía fotovoltaica en el Perú”. Tech. Rep., Servicio Nacional de Meteorología e Hidrología del Perú (2003).
48. Sánchez Ccoyllo, O. *et al.* Evaluación de la calidad del aire en lima metropolitana 2012. Tech. Rep., Servicio Nacional de Meteorología e Hidrología del Perú (2014).
49. Govindasamy, T. R. & Chetty, N. Machine learning models to quantify the influence of pm10 aerosol concentration on global solar radiation prediction in South Africa. *Clean. Eng. Technol.* **2**, 100042. <https://doi.org/10.1016/j.clet.2021.100042> (2021).
50. Cho, J., Kim, H. & Chung, Y. Spatio-temporal changes of pm10 trends in South Korea caused by East Asian atmospheric variability. *Air Qual. Atmos. Health* **14**, 1001–1016. <https://doi.org/10.1007/s11869-021-00995-y> (2021).
51. Encalada-Malca, A. A., Cochachi-Bustamante, J. D., Rodríguez, P. C., Salas, R. & López-Gonzales, J. L. A spatio-temporal visualization approach of pm10 concentration data in metropolitan Lima. *Atmosphere* **12**, 609. <https://doi.org/10.3390/atmos12050609> (2021).
52. Vardoulakis, S. & Kassomenos, P. Sources and factors affecting pm10 levels in two European cities: Implications for local air quality management. *Atmos. Environ.* **42**, 3949–3963. <https://doi.org/10.1016/j.atmosenv.2006.12.021> (2008).
53. Plocoste, T. & Pavón-Domínguez, P. Temporal scaling study of particulate matter (pm10) and solar radiation influences on air temperature in the Caribbean basin using a 3d joint multifractal analysis. *Atmos. Environ.* **222**, 117115. <https://doi.org/10.1016/j.atmosenv.2019.117115> (2020).
54. Hirth, S., Weiß, D., Nicolai, A. & Grunewald, J. Validation of models for the calculation of sun positions and mapped radiation on inclined surfaces. In *E3S Web of Conferences*, vol. 172, 03007. <https://doi.org/10.1051/e3sconf/202017203007> (EDP Sciences, 2020).
55. Gocheva-Ilieva, S., Ivanov, A. & Stoimenova-Minova, M. Prediction of daily mean pm10 concentrations using random forest, cart ensemble and bagging stacked by mars. *Sustainability* **14**, 798. <https://doi.org/10.3390/su14020798> (2022).
56. SENAMHI. Vigilancia de la calidad del aire. Tech. Rep., Servicio Nacional de Meteorología e Hidrología del Perú (2019).
57. Morton-Bermea, O. *et al.* Historical trends of metals concentration in pm10 collected in the Mexico city metropolitan area between 2004 and 2014. *Environ. Geochem. Health* **43**, 2781–2798. <https://doi.org/10.1007/s10653-021-00838-w> (2021).
58. Silva, J. *et al.* Particulate matter levels in a South American megacity: The metropolitan area of Lima-Callao, Peru. *Environ. Monit. Assess.* **189**, 635. <https://doi.org/10.1007/s10661-017-6327-2> (2017).
59. Reátegui-Romero, W. *et al.* Behavior of the average concentrations as well as their pm10 and pm2.5 variability in the metropolitan area of Lima, Peru: Case study February and July 2016. *Int. J. Environ. Sci. Dev.* **12**, 204–213 (2021).
60. Ibacache-Pulgar, G., Villegas, C., López-Gonzales, J. L. & Moraga, M. Influence measures in nonparametric regression model with symmetric random errors. *Stat. Methods Appl.* 1–25 (2022).

## Acknowledgements

The authors would like to thank the *Dirección General de Salud Ambiental* of the *Ministerio de Salud* of Peru for providing the data used in this study. They also thank *Investiga UCV* of the *Universidad César Vallejo* for financial support.

## Author contributions

All authors participated in the conceptualization, methodology, software, and manuscript writing.

## Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to R.J.C.-T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022