

Pollutant Time Series Analysis for Improving Air-Quality in Smart Cities

Raúl López-Blanco^{1*}, Miguel Chaveinte García², Ricardo S. Alonso^{2,3*}, Javier Prieto^{1*}, Juan M. Corchado^{1,2}

¹ Universidad de Salamanca, Salamanca (Spain)

² AIR Institute, Valladolid (Spain)

³ Universidad Internacional de La Rioja, Logroño (Spain)

Received 31 March 2023 | Accepted 24 July 2023 | Published 27 August 2023



ABSTRACT

The evolution towards Smart Cities is the process that many urban centers are following in their quest for efficiency, resource optimization and sustainable growth. This step forward in the continuous improvement of cities is closely linked to the quality of life they want to offer their citizens. One of the key issues that can have the greatest impact on the quality of life of all city dwellers is the quality of the air they breathe, which can lead to illnesses caused by pollutants in the air. The application of new technologies, such as the Internet of Things, Big Data and Artificial Intelligence, makes it possible to obtain increasingly abundant and accurate data on what is happening in cities, providing more information to take informed action based on scientific data. This article studies the evolution of pollutants in the main cities of Castilla y León, using Generative Additive Models (GAM), which have proven to be the most efficient for making predictions with detailed historical data and which have very strong seasonalities. The results of this study conclude that during the COVID-19 pandemic containment period, there was an overall reduction in the concentration of pollutants.

KEYWORDS

Air Pollutants, Air Quality, Climate Change, Machine Learning, Public Health.

DOI: 10.9781/ijimai.2023.08.005

I. INTRODUCTION

THE move towards Smart Cities is the evolution to which cities are tending, as they have become centres of population concentration that seek to maintain the quality of life of all their inhabitants. These increasingly overpopulated population centres. In fact, since 2008 and worldwide, there are more inhabitants in cities than in rural areas, and the trend continues to rise for cities (Fig. 1) [1].

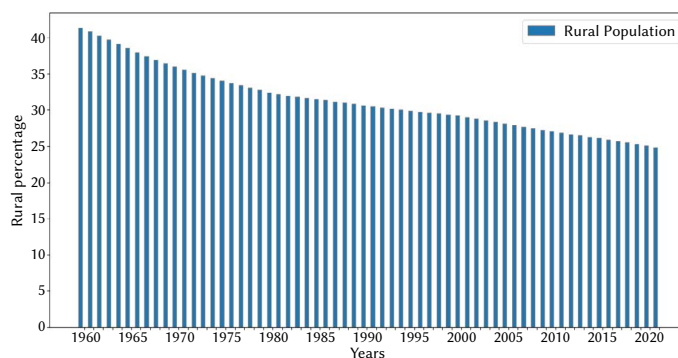


Fig. 1. Evolution of the rural population in Spain from 1960 to 2021. Source: World Bank [1].

* Corresponding author.

E-mail addresses: raullb@usal.es (R. López-Blanco), ralonso@airinstitute.com (R. S. Alonso), javierp@usal.es (J. Prieto).

People move to urban centres for many reasons, including to improve their quality of life, but maintaining the social indicators that people expect when they move to cities can be a difficult task for local, regional and other authorities involved in urban development [2]. This is partly due to some endemic problems in cities, such as air pollution, traffic (which costs €270 billion a year in Europe [3]), or the lack of green spaces (whose health benefits have been demonstrated in numerous studies [4]).

The maintenance of air quality in cities is one of the fundamental elements for the preservation and improvement of the quality of life of citizens. In fact, the World Health Organization (WHO) has a database by country that identifies the number of deaths attributable to pollution-related diseases. The WHO itself has determined that 99% of the population lives in places where the limits for pollutants suspended in the air are exceeded [5]. In Spain, for the year 2019, the estimated average percentage of deaths due to pollution was 3.32%.

A polluted environment can also influence the spread of respiratory diseases, with airborne particles acting as vectors of transmission [6] or even weakening the most vulnerable people, making them more susceptible to respiratory diseases.

At this point is where smart cities appear, seeking to improve the quality of life of citizens through the use of new technologies to achieve greater efficiency and sustainability of population services [7]. One of the main principles sought by the so-called Smart Cities is sustainability through the reduction of the environmental impact of the processes carried out in cities and the implementation of green technologies [8].

The success of the improvements introduced by Smart Cities consists of a balance between the quality of life perceived by citizens

(for example, through the introduction of green areas near residential spaces [9]), the continuous actions carried out to obtain information about the environment [10] and to know which are the critical points on which action should be taken to maintain the citizens' perception of quality of life, as well as to avoid situations of eco-anxiety [11] and other disorders derived from climate change.

The set of technologies used to collect data from cities and to have more information about what is happening in them comprises a series of innovative technologies such as:

- **The Internet of Things (IoT).** It allows to monitor the environment with different devices capable of capturing information from the surroundings, such as sensors to measure the pollutants present in the air and other magnitudes such as humidity, temperature, pressure [12].
- **Big Data.** Dealing with all the data produced by IoT devices requires a range of techniques to process and store it in the best way for later use [13].
- **Artificial Intelligence (AI).** This discipline and its most important branches such as machine learning make it possible to create predictive models from data sets [14].
- **Blockchain.** Distributed ledger technologies such as blockchain are used in smart cities to improve the efficiency, transparency and security of data management systems and services [15].

The current work mainly combines: the Internet of Things (IoT), which are those devices or stations installed in cities and responsible for capturing data on pollutants present in the environment; Big Data, which compiles all the information obtained and makes it available to researchers to carry out this type of study; as well as Artificial Intelligence, which allows modelling what happens in the environment according to variations in the data.

All the data processing has been carried out using generative additive models (GAM) that have shown better performance than other machine learning models, such as Long Short Term Memory (LSTM) networks and Autoregressive Integrated Moving Average (ARIMA) models (used as predictive models in other works that point out that predictions with this type of networks can be improved [16]).

The rest of the article is structured as follows: Section II contains a series of related works that have carried out studies on air quality in cities and that use Artificial Intelligence models to carry them out. Section III performs a predictive and evolutionary analysis of the different pollutants found in suspension in some of the most important cities of the region of Castilla y León (Spain). Section IV gathers the most important conclusions drawn from the study of the evolution of these pollutants. Finally, Section V contains the future lines of work along which the present study could advance.

II. RELATED WORKS

This section reviews some of the most important works related to the study of air quality in different urban areas. This section deals with those works that have studied the effect of airborne pollutants in the environment by different methods and how they influence the quality of life in cities [17].

In most of the occasions, more than knowing the current situation of air quality in which a city is, it is sought through historical series [18] to know what will be the evolution in the future. All this in order to know whether the trend is upward or downward for each of the pollutants and to determine whether the corrective measures that can be applied have the expected effect.

These prediction actions are carried out by means of Machine Learning models that allow modeling the behavior of the evolution

of these pollutants. Some of the most outstanding studies on pollutant evolution have been carried out using Long Short Term Memory (LSTM) networks and ARIMA models [19] and, more recently, generative additive models (GAM) [20]. Among the studies that use this type of models to predict the evolution of pollutants, those of Hasnain [21] and Shen [22] that study the evolution of pollutants in relevant cities of the Asian continent such as Seoul or regions such as Jiangsu in China, stand out.

Another area of interest on pollutants in cities is urban heat islands, areas where the temperature is significantly higher than the surrounding areas due to heat absorption and retention such as buildings and roads. Studies such as Swamy's [23] or Ngarambe's [24] have shown that heat islands can increase air pollution levels by increasing atmospheric stability and decreasing the height of the boundary layer, which limits the dispersion of pollutants.

Also important for pollutant dispersion are wind gusts, which can influence air quality by dispersing pollutants or transporting them to other areas. Studies have shown that wind gusts can influence the dispersion of fine particles in the atmosphere [25]. In addition, the presence of wind gusts can also influence the formation of pollution clouds, which can increase the levels of ozone and other pollutants in the air [26].

Air quality is directly dependent on human actions such as road traffic and industry. In fact, some studies have shown that during times of home confinement during the COVID-19 pandemic, air quality improved as virtually all air pollutants decreased [27].

From the studies reviewed, it is determined that air quality is not something that depends only on the pollutants themselves that are present in the air, but that their dispersion and concentration can be influenced by air gusts or be part of the heat island effect. The presence of these pollutants can be modeled by different Machine Learning models, being more accurate those that handle the concept of seasonality such as GAM models.

III. EXPERIMENT AND RESULTS

The experiment was based in the expansion of the task already proposed by López-Blanco et al. [20], where it was proved that the model based in the implementation of Generative Additive Model obtained better prediction results of pollutants than those obtained by LSTM and ARIMA.

To exemplify this statement, Fig. 2 and Fig. 3 demonstrate the application of LSTM recurrent networks. The main characteristic is that information can persist in the layers of the network, generating loops that allow the recall of previous states, thus creating long-term memory, which makes them ideal for learning from situations and making predictions. However, it requires data with a highly pronounced seasonality [28], which is not present in the current dataset.

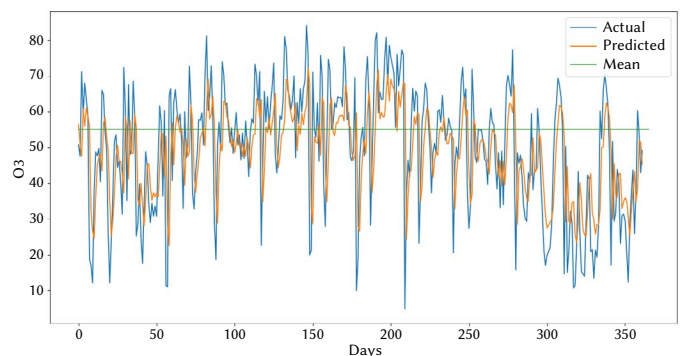


Fig. 2. Evaluation of LSTM network of pollutant O_3 in Valladolid (Spain).

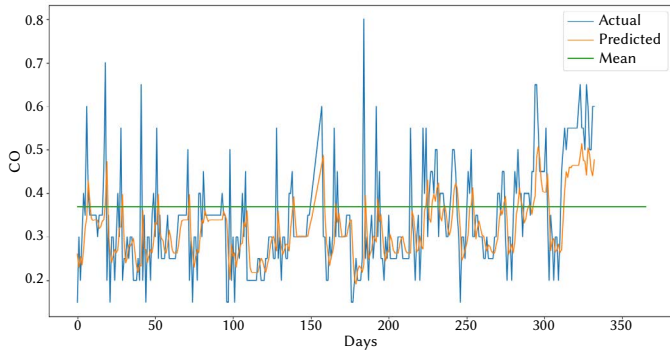


Fig. 3. Evaluation of LSTM network of pollutant CO in Valladolid (Spain).

As observed in these images, the obtained results from their evaluation do not provide predictive capability, as the networks either suffer from overfitting or impute the value of the previous day based on the considered time window.

Hence, this proposal employed the suitability of GAM as a criterion and applied it to the most populous urban areas in Castilla y León, namely: Ávila, Burgos, León, Palencia, Ponferrada, Salamanca, Segovia, Soria, Valladolid and Zamora.

Due to the previous analysis, a possible effect of the lockdown on air quality has been detected. Therefore, the spatiotemporal impact of COVID-19 lockdown measures have been evaluated in these population centers, to establish a comparison and determine the variation in atmospheric pollutant concentrations from the three years prior to the lockdown period.

A. Analysis and Forecasting Model

1. Description of the Dataset

The pollutants used in the study are: CO, NO₂, O₃ and PM_{2.5}, PM₁₀ has also been taken into account, either as a predictor or indicator of particles, in those provinces where PM_{2.5} had missing values. The CO pollutant is measured in mg/m³, while the rest are measured in µg/m³.

The presence of these pollutants in the air is a problem for human health, as many respiratory diseases have been shown to be caused by air pollution. Cancer of the respiratory tract is one of them caused in part by the presence of airborne PM [29].

All the pollutants studied affect human health, for example carbon monoxide (CO), produced by incomplete combustion of fossil fuels, reduces the blood's ability to carry oxygen; NO₂ and O₃ can cause airway irritation, respiratory problems and aggravation of asthma, hence the decision to include them in the study.

The dataset used contains daily concentration data of pollutants recorded at the air quality control stations of the Regional Government of Castilla y León [30]. The period of this data ranges from 1997 to 2020 (both included), during which there are certain periods of missing values in the different population centers studied. In general, PM_{2.5} and CO pollutants have large temporal gaps without data in most provinces, leading to various situations, which has led us to analyze each population center separately to examine their data and possible correlations between pollutants.

These facts can be observed in Fig. 4 and Fig. 5, showing the mentioned temporal evolution in the population centers of León and Ponferrada, respectively.

2. Proposed Model

For the analysis of the temporal series taking tendencies, seasonality and holidays into account, the Prophet package was used. Prophet is a tool to carry out precise and efficacious predictions, with a time

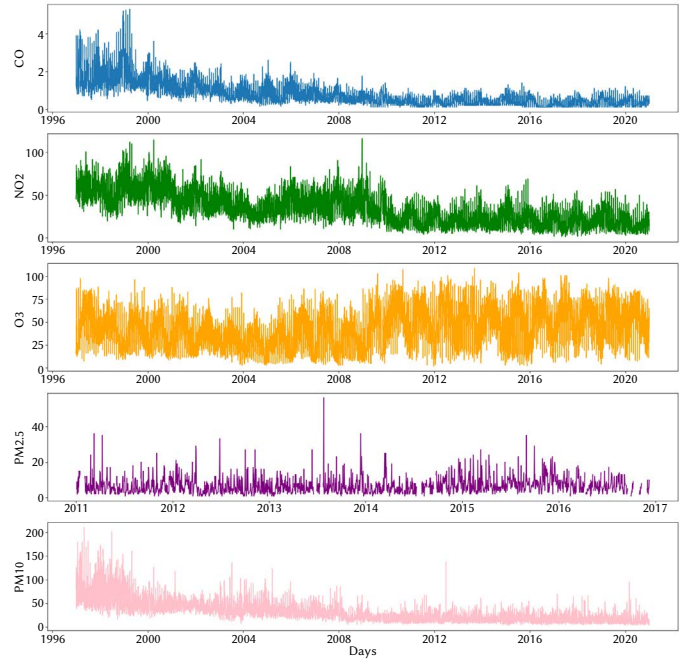


Fig. 4. Historical evolution of pollutants in the population center of León (Spain).

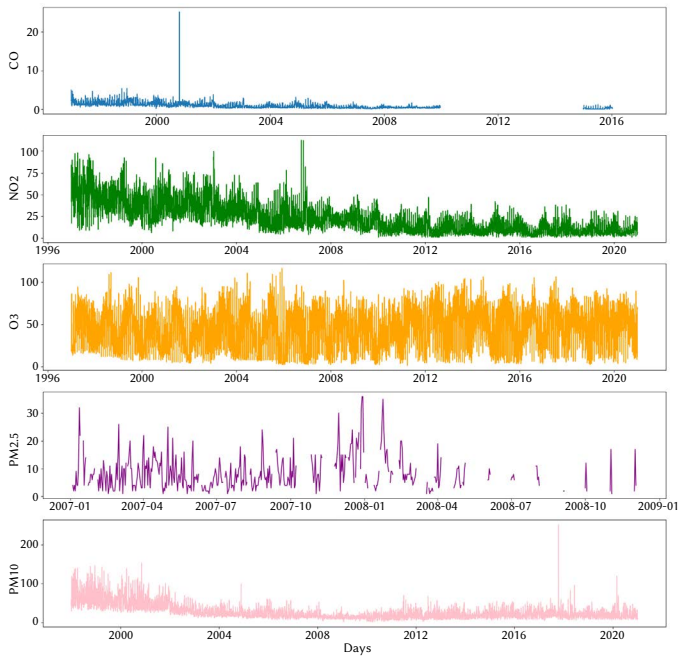


Fig. 5. Historical evolution of pollutants in the population center of Ponferrada (Spain).

of seconds to adjust the model. Equation (1) shows the expression followed by the model.

$$y(t) = g(t) + s(t) + h(t) + e(t) \tag{1}$$

The assessment of the Prophet model's performance uses (1), where $y(t)$ is the predicted value determined by a linear or logistic equation; $g(t)$, as can be seen in (1) represents non-periodic changes; seasonality is given by $s(t)$, which represents periodic changes (weekly, monthly, annual); the $h(t)$ component contributes with information about holidays and events; and finally, $e(t)$ covers the noise portion of the time series, indicating random fluctuations that cannot be predicted [31].

This results in a model composed of three sub-models: the Trend model, Seasonality model, and Holidays model [32].

The trend model, called Nonlinear Saturating Growth, is represented by the logistic growth model expressed in (2).

$$g(t) = \frac{C}{1 + \exp(-k(t - m))} \quad (2)$$

where C is the maximum capacity (the maximum value of the curve), k is the growth rate (representing the “slope” of the curve), and m is an offset parameter.

The seasonality model employs Fourier series for approximations, based on (3). The seasonal component $s(t)$ provides a flexible model of periodic changes due to weekly and annual seasonality.

$$s(t) = \sum_{n=1}^N \left(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right) \quad (3)$$

Seasonality is key when predicting new values. Prophet offers components to plot the seasonality in weekly, annual or trendy intervals taking into account the historic series. The Fig. 6 shows the seasonality plot of the O3 concentration in the population center of Valladolid in historic, annual and weekly trends. The data of the first 10 years (2011-2020) comprised the initial training data. To validate the model, we used the data from 2020 as a test for the model and were able to compare and analyze the predicted values generated by the model with the actual values. In this process, we utilized wind velocity as a regressor variable. Finally, Prophet was used to predict the air quality for the different pollutants in the years 2021 and 2022.

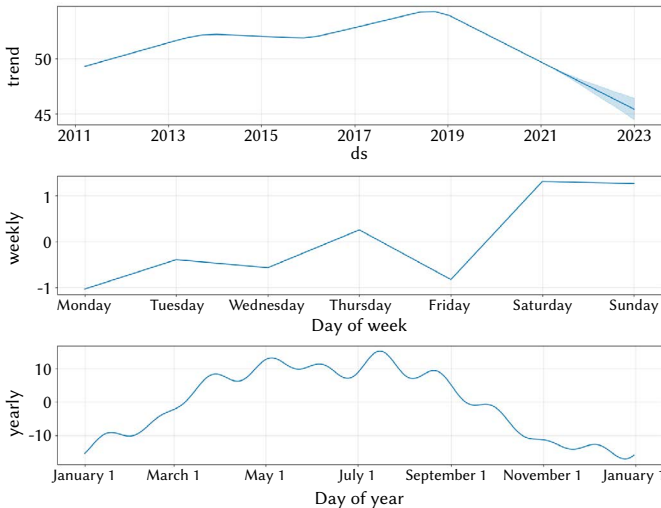


Fig. 6. Components of the Model of Valladolid (O3 trends graph, overall trend, yearly and weekly).

Therefore, this study applied GAM networks implemented in Prophet to forecast air quality. The air quality data from the Regional Government of Castilla y León [30] and the meteorological data from AEMET [33] were used as inputs. These data were preprocessed to deal with errors and missing values, using interpolation or other highly correlated pollutants as regressors. Then, the Prophet model was trained with these data to predict the values for the year 2020, incorporating wind velocity as a regressor. The predicted values were compared with the actual values, and a statistical analysis of the model performance, trend and seasonality was performed.

Finally, a two-year forecast for each pollutant was generated, examining the trends and seasonality patterns. The proposed model can be observed in Fig. 7, and its implementation in each population center and pollutant is described in Section III-A-4.

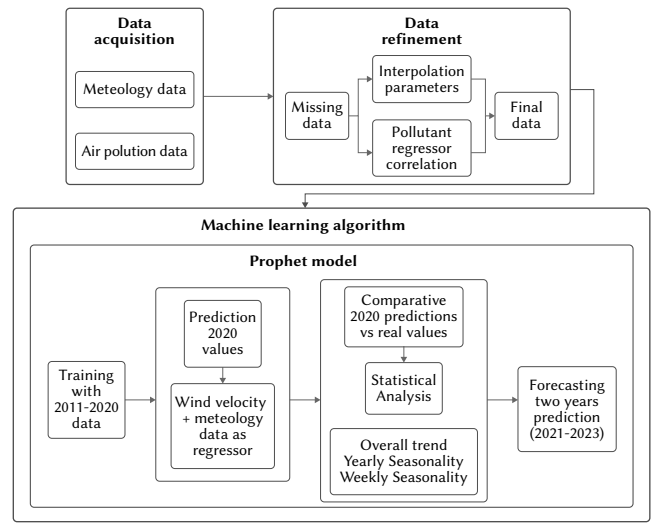


Fig. 7. The proposed architecture of prediction model for air quality.

3. Statistical Analysis

In assessing the model’s efficacy, various statistical measures were computed: Pearson’s correlation coefficient, mean squared error, root mean squared error, and mean absolute error. The Pearson correlation coefficient (R) was employed to ascertain if the model exhibited overfitting or underfitting. Optimal values are approximately 0.5, indicating that the model adheres to the series’ overall pattern without overfitting.

The mean squared error (MSE) represents the average squared discrepancy between estimated and predicted values. The root mean squared error ($RMSE$) is the square root of MSE . The mean absolute error (MAE) is determined by averaging the differences between given and actual values. As $RMSE$ assigns greater weight to outliers compared to MAE , the disparity between the two reflects the influence of outliers within the dataset [21].

$$R = \frac{n \sum_{i=1}^n (x_i \hat{x}_i) - (\sum_{i=1}^n x_i)(\sum_{i=1}^n \hat{x}_i)}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n \hat{x}_i^2 - (\sum_{i=1}^n \hat{x}_i)^2}} \quad (4)$$

$$MSE = \sum_{i=1}^n (x_i - \hat{x}_i)^2 \quad (5)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i| \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2} \quad (7)$$

where x_i and \hat{x}_i are the actual and predicted values respectively, and n represent the number of samples.

4. Results and Discussion

After selecting the algorithm to be employed, it is necessary to mention that it has been decided to implement the Generative Additive Model (GAM) from Prophet, which results highly convenient for the data series which encompass extense periods of detailed historic observations, with pronounced seasonalities which involve previously identified relevant, although irregular elements, as well as data points with significative outliers whose non linear growth trends approach a limit.

The prediction of the temporal series can be observed in Fig. 8, along with the seasonality in Fig. 6, and later in the detailed analysis which is explained after each pollutant. To carry out this prediction and due to the casuistics of the previously commented data, it has been opted to use the NO_2 as additional regressor to predict the missing values in the $\text{PM}_{2.5}$ and CO series that had a strong linear correlation (Pearson correlation coefficient). This method has been carried out in the population center of Salamanca (0.71 and 0.59, respectively) and León (0.54 and 0.73, respectively).

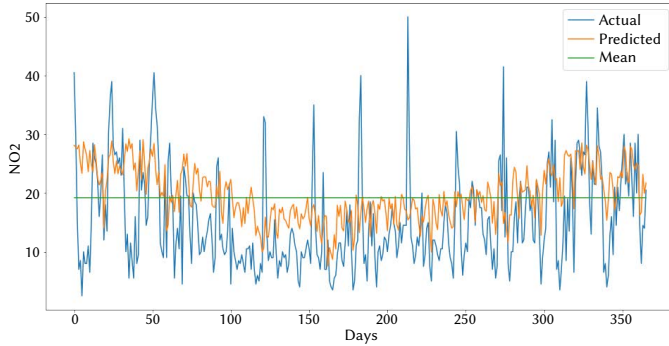


Fig. 8. NO_2 ($\mu\text{g}/\text{m}^3$) forecasting vs. real values 2020 in León (Spain).

In Valladolid it was only used for the pollutant $\text{PM}_{2.5}$ (0.65).

In Burgos and Ponferrada, the PM_{10} was previously used as regressor for the $\text{PM}_{2.5}$ due to its high correlation (0.75 and 0.88 respectively). Subsequently in Burgos (0.77 and 0.76 respectively) the previous method was used for the prediction of the $\text{PM}_{2.5}$ and CO.

In Soria, Zamora, Palencia and Segovia the same method is used, but working with the PM_{10} since there is not enough data to predict the $\text{PM}_{2.5}$.

In Ávila, for the CO pollutant there are neither data or ways to correlate them to make a prediction. And for the particle analysis the PM_{10} was used due to the $\text{PM}_{2.5}$ not having enough data.

In constructing the models, various seasonalities have been employed: weekly and annual. This is done to account for the impact of predefined Spanish holidays and the influence of weekly traffic patterns, which are higher on weekdays and lower on weekends. With these considerations, a one-year prediction is made; and for this purpose, from the initial dataset, we retain the part of the historical series with the least noise and the longest length possible. Thus, we truncate the data's beginning, as seen in Fig. 4 and Fig. 5, which exhibits more noise and has a higher value difference compared to more recent data. Consequently, the first 14 years for NO_2 and O_3 and the first 5 years for CO and particle analysis, whether

$\text{PM}_{2.5}$ or PM_{10} in provinces, have been removed due to the aforementioned casuistry.

Within each pollutant, the followed seasonality will be examined in detail. To verify the goodness of the model's performance, the statistics mentioned earlier in Section 3 have been utilized.

NO_2

The results of the analysis are shown in Table I, where the Pearson correlation coefficient, the *MSE*, the *RMSE* and the *MAE* can be seen. These values denote the appropriateness of the model in fitting the historical data, as well as the accuracy of predictions; for instance, Salamanca exhibits both the highest Pearson correlation coefficient (0.68) and the lowest RMSE ($3.46\mu\text{g}/\text{m}^3$) which implies that its model best catches the overall trend and the lowest forecast error. On the other hand, Soria has the lowest Pearson correlation coefficient (0.37) as well as the highest RMSE ($10.12\mu\text{g}/\text{m}^3$), which means that its model has the worst fit to the general trend and the highest forecast error as

well as a wide confidence interval. This results might be due to Soria's series having more noise, more variability or more external factors affecting its behavior.

TABLE I. NO_2 2020 MODEL PERFORMANCE STATISTICS FOR THE DIFFERENT POPULATION CENTERS

Pop. Center	R	MSE	RMSE	MAE
Ávila	0.43	19.33	4.40	3.13
Burgos	0.37	38.56	6.21	4.66
León	0.48	72.13	8.49	6.93
Palencia	0.38	28.33	5.32	4.16
Ponferrada	0.48	19.51	4.42	3.03
Salamanca	0.68	11.98	3.46	2.74
Segovia	0.36	23.92	4.89	3.72
Soria	0.37	102.40	10.12	8.45
Valladolid	0.41	58.91	7.68	6.19
Zamora	0.47	31.84	5.64	4.22

In addition to the statistical data from conducting the proposed analysis, it has been decided to plot a comparison between the actual values and the predicted values in the year 2020 and analyze the behavior of the model visually.

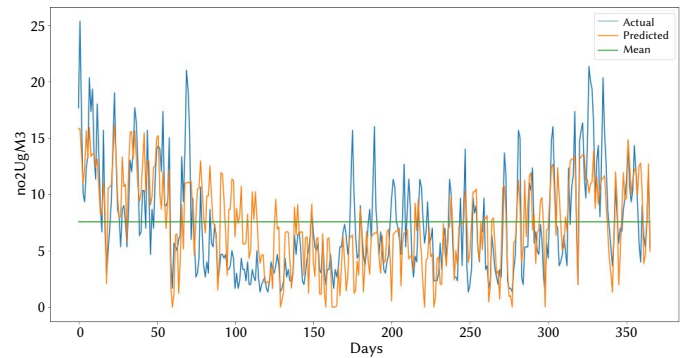


Fig. 9. NO_2 ($\mu\text{g}/\text{m}^3$) forecasting vs. real values 2020 in Salamanca (Spain).

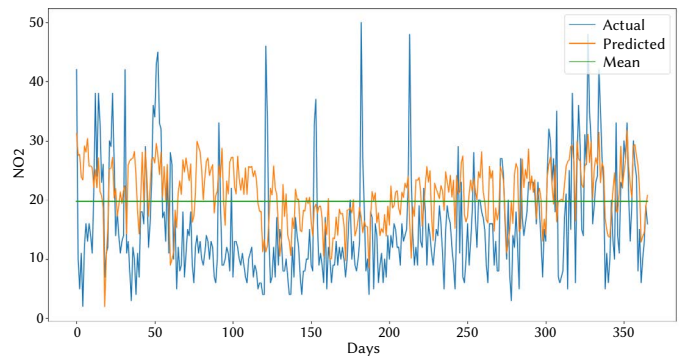


Fig. 10. NO_2 ($\mu\text{g}/\text{m}^3$) forecasting vs. real values 2020 in Soria (Spain).

We can see in Fig. 8, Fig. 9 and Fig. 10 that the adjustment in the examples is good, even predicting peaks in the series; which confirms the statistical values of this pollutant discussed. We can also observe that between the periods of March and May 2020, the predicted values remain above the actual values, which is a general fact in all the analyzed population centers.

Finally, we performed a two-year prediction in which the prediction is displayed alongside the error margins, where the black points represent the actual values, the dark blue trendline is the temporal pattern that the model learns from and uses for predictions, and the lighter blue areas represent the error margins of the two-year prediction, where the actual values are no longer shown.

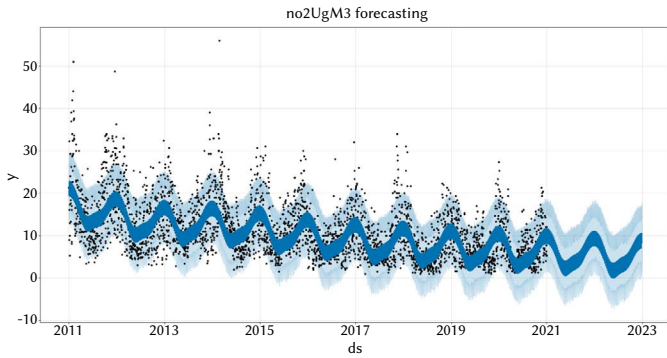


Fig. 11. NO_2 ($\mu\text{g}/\text{m}^3$) two years prediction in Salamanca (Spain).

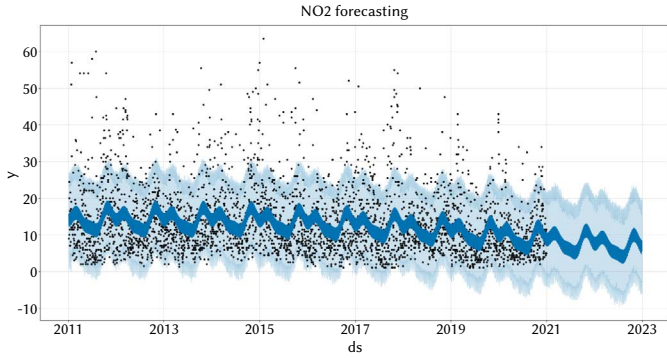


Fig. 12. NO_2 ($\mu\text{g}/\text{m}^3$) two years prediction in Burgos (Spain).

In Fig. 11 and Fig. 12 we can verify how the statistical results translate into the prediction, the trend of the series and the possible outliers, and their effect on the model.

Along with this, we obtain the components of the model, in which seasonality plays a prominent role, as discussed in the model (2).

In all the analyzed population centers, the trend observed in Fig. 13 is followed, which is exemplified by Valladolid. In this figure, we can see how the trend in recent years for NO_2 concentration is decreasing, and the prediction is that it will continue this pattern in the coming years. Weekly, it follows a stable pattern during workdays, declining on weekends. As for the annual trend, it experiences a decrease from March to August, with the highest values occurring during the first and third four-month period of the year.

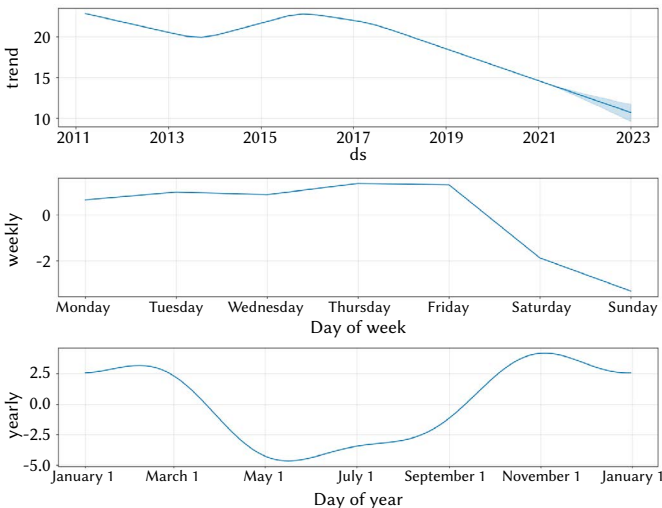


Fig. 13. Components of the Model of Valladolid (NO_2 trends graph, overall trend, yearly and weekly).

PM_{2.5}

From the results in Table II corresponding to the analysis of the $\text{PM}_{2.5}$ in the population centers where data or correlation was available, the following conclusions can be drawn:

TABLE II. $\text{PM}_{2.5}$ 2020 MODEL PERFORMANCE STATISTICS FOR THE DIFFERENT POPULATION CENTERS

Pop. Center	R	MSE	RMSE	MAE
Burgos	0.25	6.69	2.59	1.98
León	0.40	6.25	2.50	2.04
Ponferrada	0.17	89.51	9.46	5.50
Salamanca	0.30	12.45	3.53	2.65
Valladolid	0.24	23.02	4.80	3.90

The effectiveness of the prediction models varies between different population areas, as demonstrated by the R , MSE , $RMSE$, and MAE values. In general terms, the $\text{PM}_{2.5}$ prediction models show variable performance in different population areas, with a moderate fit in most cases. This suggests that the models capture the general trend in $\text{PM}_{2.5}$ pollution levels but are not overfitted. Greater prediction accuracy is observed in areas such as León, while in areas like Ponferrada, the model's performance is lower. Pearson correlation coefficients (R) range between 0.17 and 0.40, suggesting that there is some variability in the quality of the predictions between different population areas.

It is observed that some areas, such as León, have a higher Pearson correlation coefficient (0.40) and a lower error ($RMSE$ of 2.50), indicating that the prediction model is more accurate in these areas (Fig. 14). On the other hand, areas like Ponferrada show a lower correlation coefficient (0.17) and a higher error ($RMSE$ of 9.46), suggesting a lower performance of the model in this area (Fig. 15).

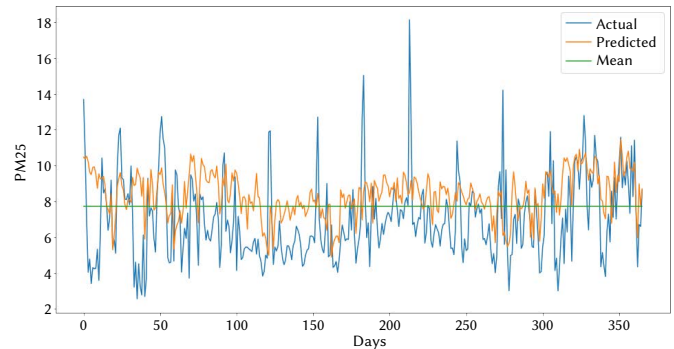


Fig. 14. $\text{PM}_{2.5}$ ($\mu\text{g}/\text{m}^3$) forecasting vs. real values 2020 in León (Spain).

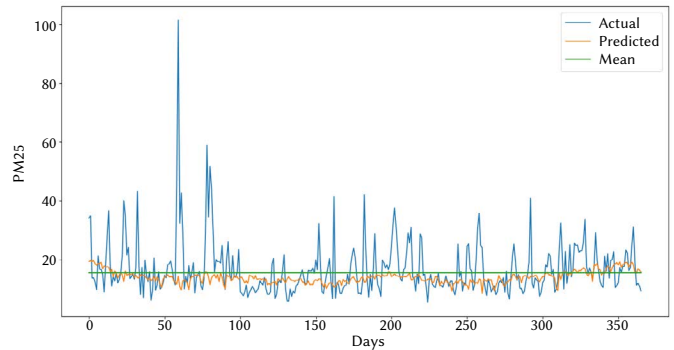


Fig. 15. $\text{PM}_{2.5}$ ($\mu\text{g}/\text{m}^3$) forecasting vs. real values 2020 in Ponferrada (Spain).

In Fig. 16 and Fig. 17, we observe the two-year predictions for these pollutants in the cases of Burgos and León respectively. We see that outliers appear, but as observed in the statistical analysis, the behavior in León is superior, adjusting to the stationary trend.

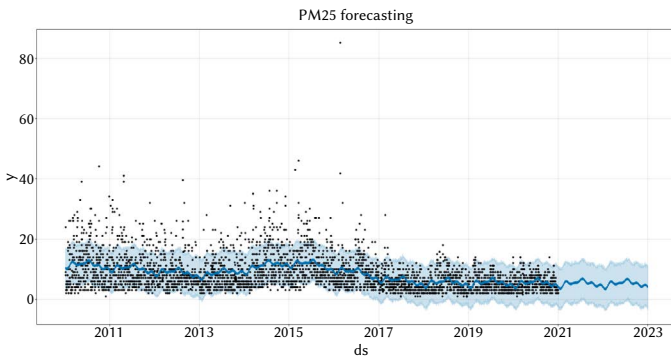


Fig. 16. $PM_{2.5}$ ($\mu g/m^3$) two years prediction in Burgos (Spain).

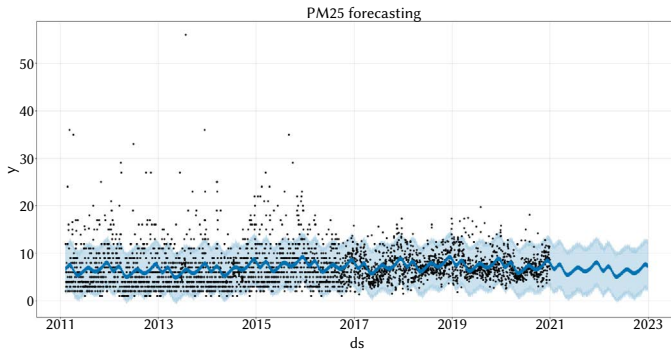


Fig. 17. $PM_{2.5}$ ($\mu g/m^3$) two years prediction in León (Spain).

Regarding the trend followed by this pollutant (Fig. 18), a decrease is observed in relation to recent years, predicting that this pattern will continue in the coming years. Weekly, it reaches its maximum peak during the middle of the week, being lower during the first and last days of the week. The same occurs inversely in the annual trend, with the first quarter and the last four months of the year being the highest points, and oscillating during the second quarter at the lowest values.

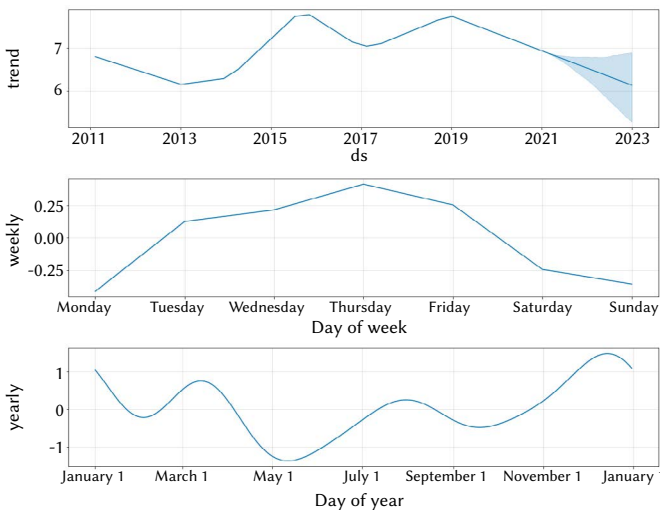


Fig. 18. Components of the Model of León ($PM_{2.5}$ trends graph, overall trend, yearly and weekly).

PM_{10}

The PM_{10} prediction models also show variable performance among the analyzed population areas (where $PM_{2.5}$ analysis was not possible), with a moderate fit in most cases. This indicates that the models capture the general trend in PM_{10} pollution levels without being overfitted. Differences in R , $RMSE$, and MAE values between population areas

suggest variability in the quality of predictions across different areas. Soria serves as an example of the best-performing case. In Soria, the PM_{10} prediction model exhibits a Pearson correlation coefficient (R) of 0.29, an $RMSE$ of 8.63, and an MAE of 5.96. These values indicate a moderate fit and acceptable performance in predicting pollution levels in this area. This contrasts with Segovia, where the PM_{10} prediction model displays a Pearson correlation coefficient (R) of 0.23, an $RMSE$ of 10.65, and an MAE of 7.33. Although the model's fit is moderate, its performance is inferior compared to the case of Soria. The lower correlation and higher error values indicate that the model may not be as accurate in predicting PM_{10} pollution levels in Segovia (Table III).

TABLE III. PM_{10} 2020 MODEL PERFORMANCE STATISTICS FOR THE DIFFERENT POPULATION CENTERS

Pop. Center	R	MSE	RMSE	MAE
Ávila	0.28	135.23	11.63	6.26
Palencia	0.27	82.71	9.09	5.99
Segovia	0.23	113.32	10.65	7.33
Soria	0.29	74.51	8.63	5.96
Zamora	0.27	66.23	8.14	5.13

PM_{10} exhibits a moderate and variable adjustment, similar to $PM_{2.5}$, depending on the analyzed population center. The model's prediction for 2020 is shown in Fig. 19 and Fig. 20, taking the cases of Soria and Segovia, respectively. We see that, like $PM_{2.5}$, the adjustment is highly sensitive to possible outliers and changes in peaks reached by the series. This can be explained by the fact that NO_2 does not work well as a regressor for filling missing values in the series in all population centers.

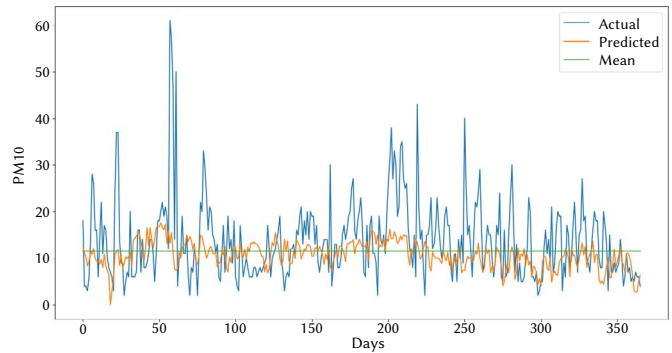


Fig. 19. PM_{10} ($\mu g/m^3$) forecasting vs. real values 2020 in Soria (Spain).

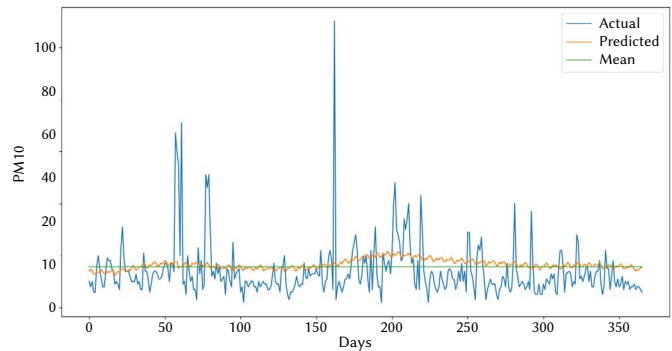


Fig. 20. PM_{10} ($\mu g/m^3$) forecasting vs. real values 2020 in Segovia (Spain).

Similarly, the two-year prediction and its adjustment to the trend and the impact of outliers are presented in Fig. 21 and Fig. 22.

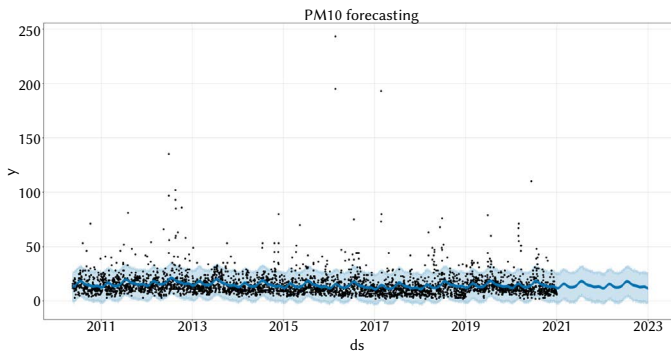


Fig. 21. PM_{10} ($\mu g/m^3$) two years prediction in Segovia (Spain).

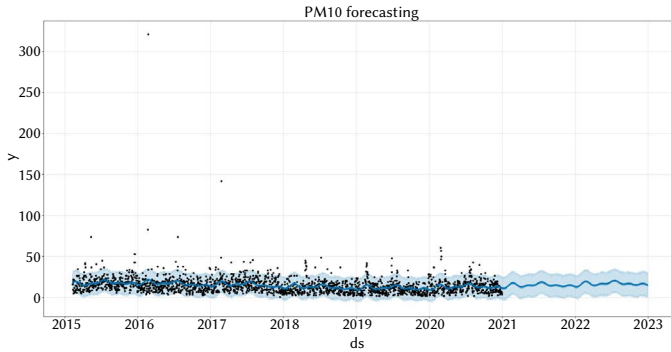


Fig. 22. PM_{10} ($\mu g/m^3$) two years prediction in Soria (Spain).

The trend followed in the urban centers where PM_{10} has been analyzed shows an upward tendency in recent years, which is expected to continue growing (Fig. 23). On a weekly basis, it follows the pattern of the highest values during workdays, decreasing to the minimum values on weekends. Annually, the stationary trend oscillates throughout the year, reaching maximums in March and in July-August, and decreasing to minimums at the end of the year.

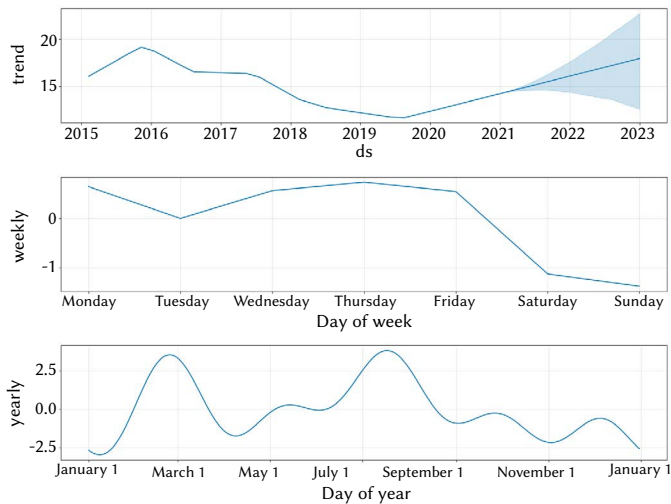


Fig. 23. Components of the Model of Soria (PM_{10} trends graph, overall trend, yearly and weekly).

O_3

The performance of time series prediction models for the O_3 pollutant varies depending on the population centers under consideration, as evidenced by the calculated evaluation measures (Table IV). It is noted that Salamanca exhibits the best model fit, with high accuracy and low variability, followed by Burgos and Segovia, which display

low-to-medium accuracy and low-to-medium variability. In contrast, Zamora presents the worst model fit, with moderate-to-high accuracy and very high variability, followed by Palencia and Ponferrada, which demonstrate moderate-to-low accuracy and high variability. The remaining population centers show intermediate values between these two groups. These differences can be attributed to various factors that influence the nature of the time series for each city, such as data quality, seasonality, the cyclical component, complexity, and heterogeneity.

TABLE IV. O_3 2020 MODEL PERFORMANCE STATISTICS FOR THE DIFFERENT POPULATION CENTERS

Pop. Center	R	MSE	RMSE	MAE
Ávila	0.45	277.98	16.67	12.69
Burgos	0.36	221.97	14.90	11.48
León	0.46	266.98	16.34	12.26
Palencia	0.43	312.71	17.68	13.32
Ponferrada	0.47	288.22	16.98	13.50
Salamanca	0.61	175.88	13.26	10.44
Segovia	0.47	231.92	15.23	11.91
Soria	0.42	214.86	14.66	11.18
Valladolid	0.49	278.60	16.91	12.69
Zamora	0.49	335.15	18.31	14.40

In Fig. 24 and Fig. 25, we can observe the model fit in the 2020 prediction alongside the actual values. Salamanca and Zamora are shown, as previously mentioned, as examples of the results of the statistical analysis. In these images, it is demonstrated that Salamanca's fit is better, even successfully predicting maximum peaks accurately.

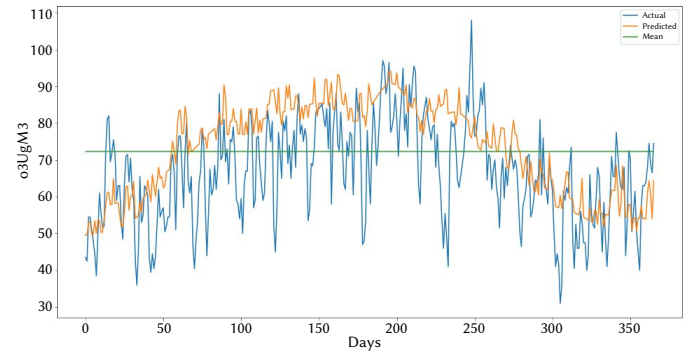


Fig. 24. O_3 ($\mu g/m^3$) forecasting vs real values 2020 in Salamanca (Spain).

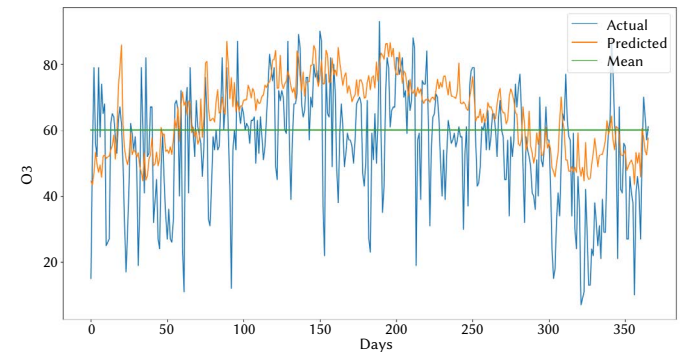


Fig. 25. O_3 ($\mu g/m^3$) forecasting vs real values 2020 in Zamora (Spain).

Additionally, the two-year predictions for the other pollutants are included. In Fig. 26 and Fig. 27, we can see in more depth the trend fit and the differences in the model fit between both provinces, and the reason for the variability detected with the higher RMSE in Zamora, due to the presence of a larger number of outliers resulting in a slightly worse prediction fit.

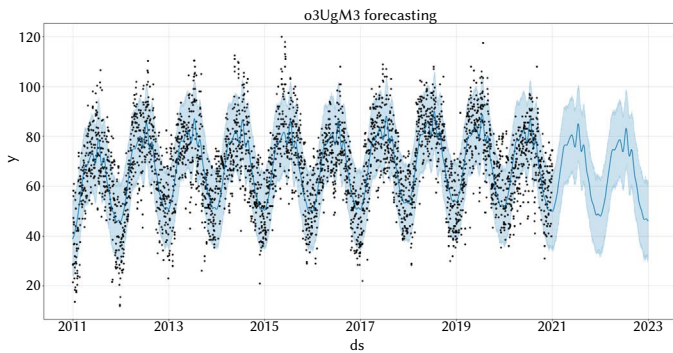


Fig. 26. O_3 ($\mu g/m^3$) two years prediction in Salamanca (Spain).

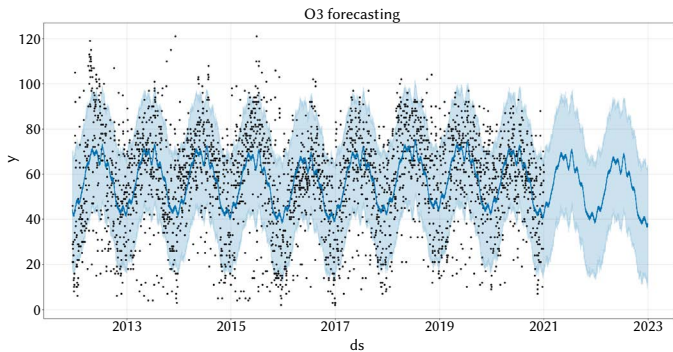


Fig. 27. O_3 ($\mu g/m^3$) two years prediction in Zamora (Spain).

Regarding the stationary trend of the O_3 pollutant, the graphs in Fig. 6 and Fig. 28 of Valladolid and Segovia urban centers are included (as previously mentioned, one urban center is used as an example of the general pattern). In these figures, it can be seen that since 2019, the trend has been decreasing and is expected to continue. Weekly, a pattern similar to that presented by NO_2 is found, with an increase on weekends. Moreover, in O_3 , there is a certain midweek peak. Annually, it is observed that the highest values are reached in the months of the second quarter of the year, while the lowest values occur in the rest of the quarters.

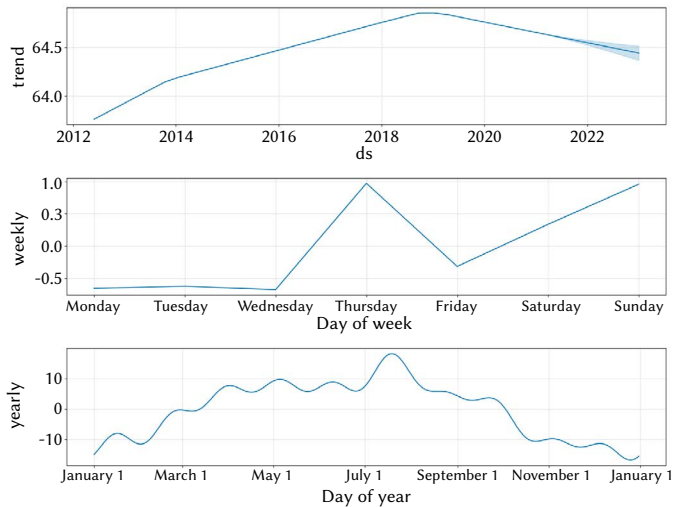


Fig. 28. Components of the Model of Segovia (O_3 trends graph, overall trend, yearly and weekly).

CO

Upon analyzing the results Table V corresponding to the CO pollutant in different population centers, we can draw several conclusions. Firstly, it can be seen that in Avila there are no adequate data or

correlations available to predict CO levels. Regarding the performance of the models in other areas, significant variations are noticed in terms of Pearson correlation coefficient (R), MSE , $RMSE$, and MAE .

TABLE V. CO 2020 MODEL PERFORMANCE STATISTICS FOR THE DIFFERENT POPULATION CENTERS

Pop. Center	R	MSE	RMSE	MAE
Ávila	-	-	-	-
Burgos	0.06	0.03	0.17	0.14
León	0.25	0.05	0.22	0.19
Palencia	0.01	0.01	0.08	0.07
Ponferrada	0.88	0.01	0.07	0.05
Salamanca	0.17	0.06	0.24	0.19
Segovia	0.09	0.02	0.13	0.11
Soria	0.28	0.005	0.07	0.05
Valladolid	0.02	0.02	0.16	0.13
Zamora	0.77	0.005	0.07	0.06

For example, in Ponferrada and Zamora, the models seem to be overfitted, as they exhibit very high Pearson correlation coefficients (0.88 and 0.77, respectively). This could be due to the use of NO_2 as a regressor for the missing CO values in their respective series.

In other areas, such as Burgos, León, Palencia, Salamanca, Segovia, Soria, and Valladolid, the results vary in terms of fit and accuracy. Some areas like León and Soria (Fig. 29 and Fig. 30) show moderate correlation coefficients (0.25 and 0.28, respectively), while others such as Palencia and Valladolid display very low correlations (0.01 and 0.02, respectively).

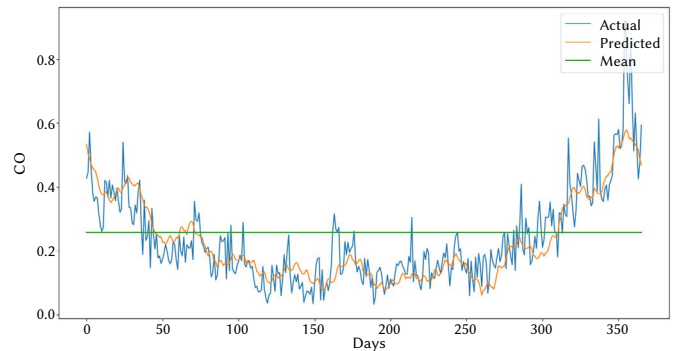


Fig. 29. CO_3 (mg/m^3) forecasting vs. real values 2020 in Ponferrada (Spain).

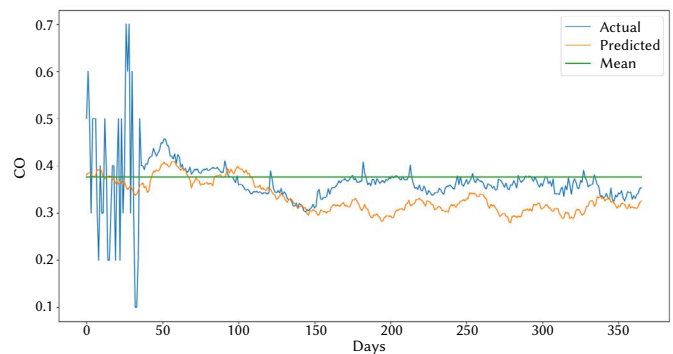


Fig. 30. CO_3 (mg/m^3) forecasting vs. real values 2020 in Soria (Spain).

This effect is even more pronounced in the two-year prediction, which partly explains the obtained statistical values. In Fig. 31, it can be seen how the model follows the series trend and is capable of approximating the periods with missing values since it has enough data and does not present a large number of outliers that might confuse it. Meanwhile, in Fig. 32, the prediction is not entirely accurate due to the large number of outliers and some periods in the series that lack of data.

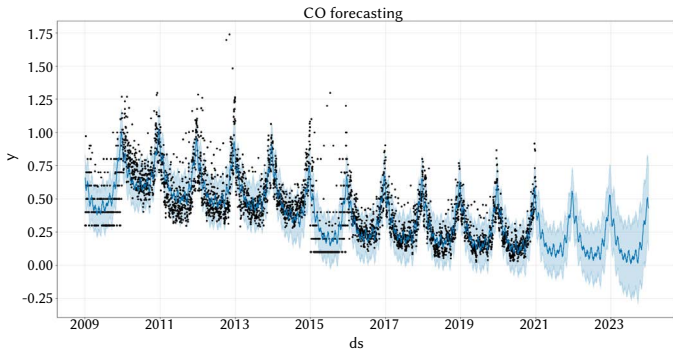


Fig. 31. CO₃ (mg/m³) two years prediction in Ponferrada (Spain).

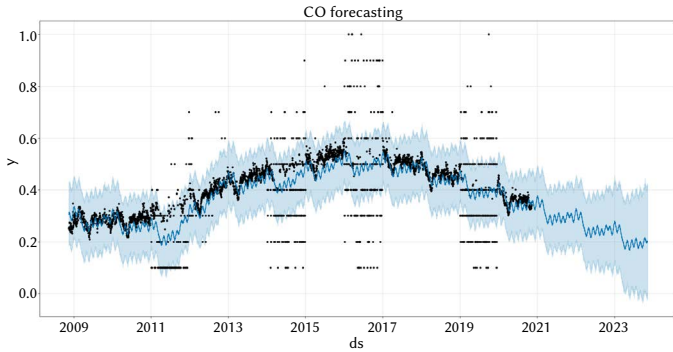


Fig. 32. CO₃ (mg/m³) two years prediction in Soria (Spain).

In summary, the table results indicate that the models used to predict CO levels in different population centers present variations in their performance and accuracy. These variations can be attributed to differences in modeling approaches, data quality, and correlations between the pollutants used as regressors. In future research, it would be advantageous to investigate alternative modeling approaches and additional factors, such as wind gusts or the so-called heat island effect, with the aim of enhancing the accuracy of CO predictions in these population areas.

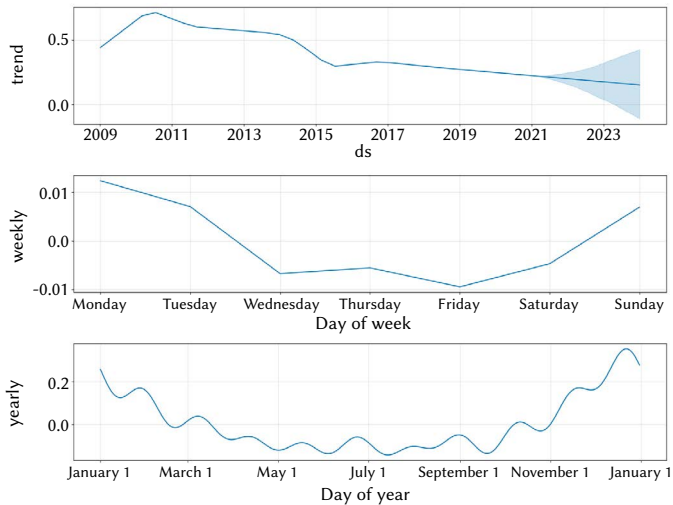


Fig. 33. Components of the Model of Ponferrada (CO trends graph, overall trend, yearly and weekly).

Finally, regarding the trend, the study of the stationary trend is presented, specifically for Ponferrada, but generalizing to the rest of the urban centers. In Fig. 33, it can be observed that there has been a downward trend in recent years, although it has slowed down and stagnated. Weekly, during the weekend, the values increase, reaching

the highest peak between Sunday and Monday, and then decreasing after Monday, taking the lowest values between Wednesday and Friday. Annually, it can be observed that the trend begins to increase from October until the end of the year. During the first two quarters of the year, the trend is decreasing.

B. COVID-19's Impact on Air Quality

As analyzed in the one-year predictions in Section 4, the predicted value of pollutants in general, although particularly notable in NO₂, is higher than the actual values between the periods of March and May 2020, while it adjusts for the rest of the year's prediction, even to the highest peaks. According to numerous studies, a sudden decrease in pollutant concentrations has been observed worldwide: Malaysia [34], northern China [35] and Brescia (Lombardy) [36].

This has led us to investigate this period in depth and how it fits within the historical time series of each population centers. To this end, in this part of the research, we partitioned the time series data into six distinct periods for analysis. The pre-lockdown phase spanned from December 1, 2019, to March 13, 2020 (103 days), while the lockdown period extended from March 14, 2020, to June 21, 2020 (99 days). The post-lockdown phase occurred between June 22, 2020, and September 30, 2020 (100 days). Additionally, we included three comparative periods (P4-P6), which corresponded to the same lockdown dates in the years 2017, 2018 and 2019.

The results are shown by pollutant with their respective spatiotemporal variations in each population centers. To perform the analysis, with the mentioned dates, they have been combined to provide a perspective on air quality during the lockdown period. The following variations (in%) were considered in averaged over the periods detailed below (the order followed is important, as it corresponds to the row number of the variation in the heatmap):

1. Variation between the lockdown period and the pre-lockdown period.
2. Variation between the post-lockdown period and the period ranging from the beginning of the pre-lockdown to the end of the lockdown.
3. Variation between the 2020 lockdown period and the average of the same dates in 2017, 2018, and 2019.
4. Variation between the average of the entire year 2020 and the average of the entire year 2019.
5. Variation between the average of the entire year 2020 and the average of the years 2017, 2018, and 2019.

NO₂

As previously discussed, one of the most notable effects of this decrease occurs in NO₂. It has been decided to display the analysis of the different proposed variations in a heatmap, as shown in Fig. 34. In

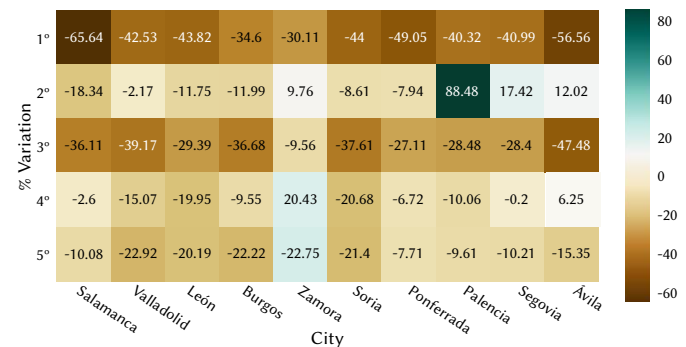


Fig. 34. Variation of NO₂ in different time periods for the different population centers.

this heatmap, it can be observed that the impact of the lockdown has led to a significant reduction in the 2020 lockdown period compared to the average of previous years in all studied population centers, in more detail in Fig. 35. This has resulted in a generalized decrease in this pollutant in 2020 compared to previous years.

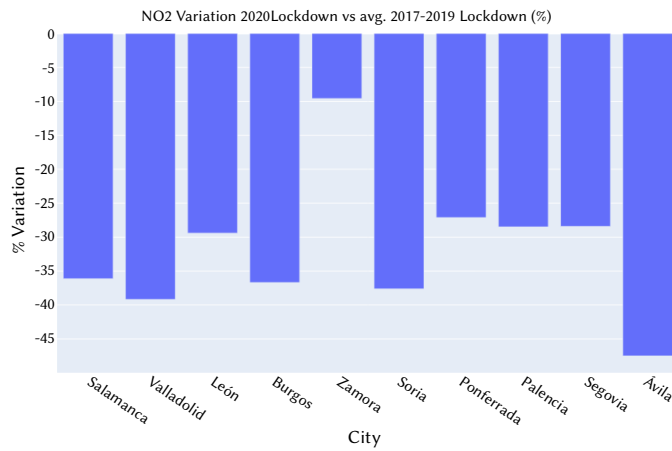


Fig. 35. Variation of NO₂ between the 2020 lockdown period and the average of the same dates in 2017, 2018, and 2019 for the different population centers.

The largest percentage decreases are found in the first of the proposed variations. In this case, in addition to the "lockdown factor", the decrease in values is due to the stationary trend of the pollutant (Fig. 13) that occurs during the lockdown period.

O₃

In the case of the O₃ pollutant, a similar behavior is observed during the lockdown period, as shown in Fig. 36, although its decrease is not as pronounced.

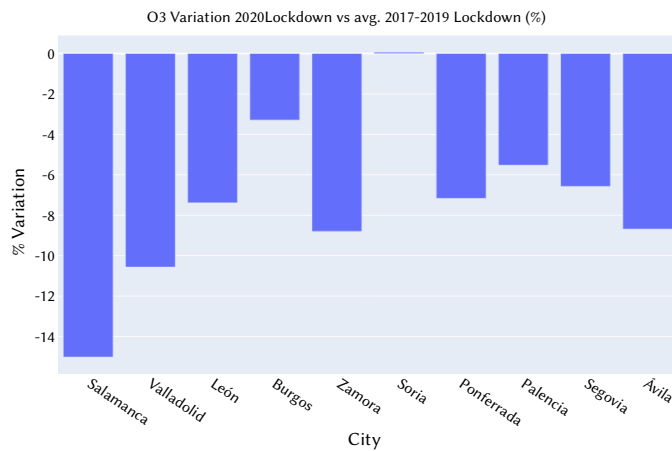


Fig. 36. Variation of O₃ between the 2020 lockdown period and the average of the same dates in 2017, 2018, and 2019 for the different population centers.

At the same time, as can be seen in the heatmap of this pollutant in Fig. 37, this leads to a decrease in the average values in 2020 across all population centers compared to the other years analyzed.

The increase observed during the lockdown period, compared to the period preceding it, is due to the stationary trend in which the highest peaks are reached during the lockdown, as can be seen in Fig. 28.

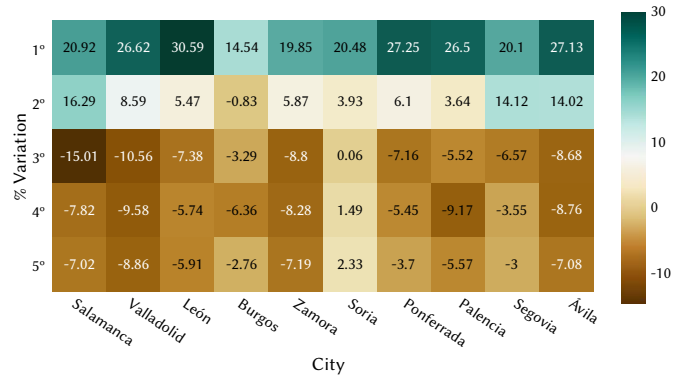


Fig. 37. Variation of O₃ in different time periods for the different population centers.

PM_{2.5}

In the population centers with data available for this pollutant, this analysis has been carried out. In this pollutant, we begin to see disparate behaviors among population centers during the lockdown period. In that period, compared to the average of previous years, only Valladolid and León experience a significant decrease (-12.47% and -10.48%, respectively) as shown in Fig. 38. Meanwhile, in the rest of the provinces, there is a slight increase in the following order: Ponferrada (+0.68%), Salamanca (+2.68%), and Burgos (+4.50%).

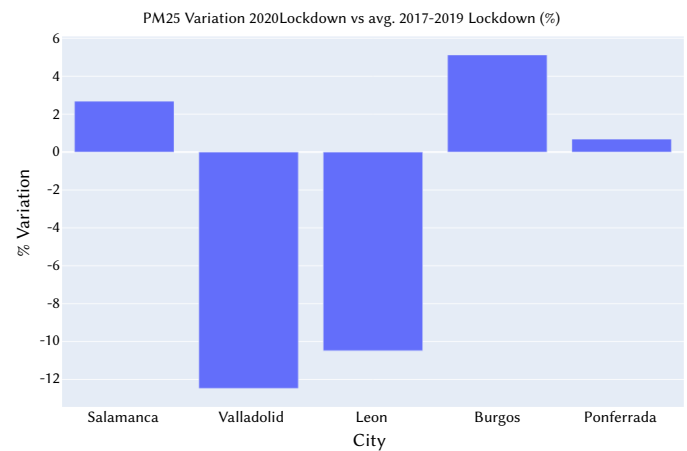


Fig. 38. Variation of PM_{2.5} between the 2020 lockdown period and the average of the same dates in 2017, 2018, and 2019 for the different population centers.

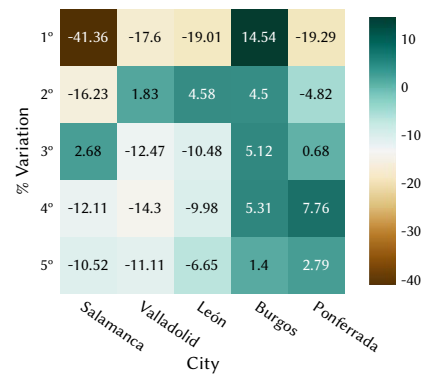


Fig. 39. Variation of PM_{2.5} in different time periods for the different population centers.

Some of the most striking data are those presented by Salamanca and Burgos concerning the variation between the lockdown period and the one immediately preceding it (Fig. 39). Furthermore, both values are at the extremes of the scales and outside the range of the other

population centers: While Salamanca shows a decrease of -41.36%, Burgos increases by 14.54% during that period. The data provided by Burgos is especially noteworthy, given that the stationary trend in that period is a decrease in pollutant values.

Finally, it should be noted that in the population centers of Salamanca, Valladolid, and León, there is a decrease in PM_{2.5} values during 2020, while in Burgos and Ponferrada, there is an increase compared to 2019 and another one during the years 2017, 2018, and 2019 although at a lower rate.

PM₁₀

The analysis has been carried out for those population centers where it was not possible to do so with the PM_{2.5} particle pollutant, due to the lack of data for that period and pollutant.

During the lockdown period, as seen in Fig. 40, all population centers manage to reduce their values compared to the average of previous years, achieving a significant decrease in Zamora (-27.93%) and Palencia (-27.46%); except for Ávila, which increased its values by 12.43% in this comparison.

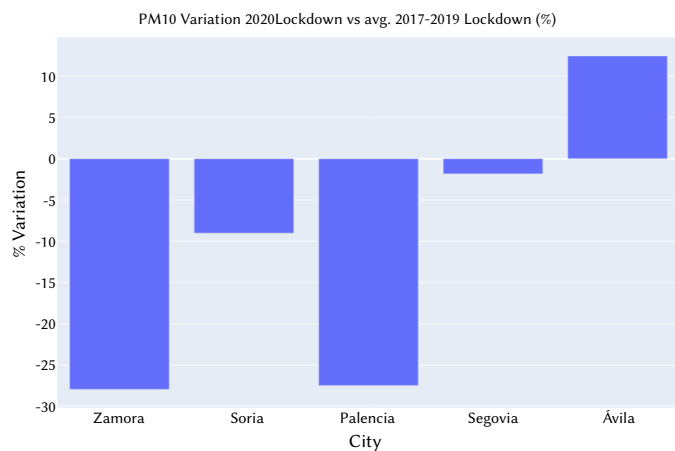


Fig. 40. Variation of PM10 between the 2020 lockdown period and the average of the same dates in 2017, 2018, and 2019 for the different population centers.

All population centers follow the stationary trend in PM₁₀, experiencing a decrease during the lockdown period compared to the previous period, as observed in Fig. 41.

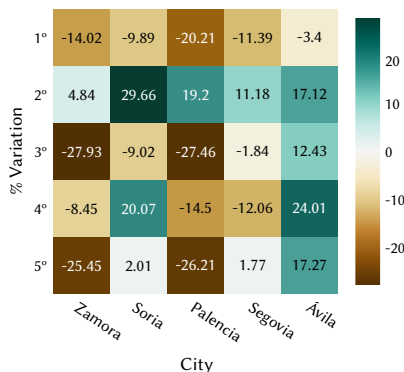


Fig. 41. Variation of PM₁₀ in different time periods for the different population centers.

When comparing the data for the entire year 2020, Ávila is again found to be one of the provinces that stood out from the rest, registering an increase of up to 24% compared to 2019 and 17.27% compared to the averages of the years 2017, 2018, and 2019. Soria also stands out in the comparison of 2020 with 2019, with an increase in data of 20.07%.

A generalized decrease is observed in the rest of the population centers.

CO

The analysis of the CO pollutant has yielded diverse results (Ávila lacked data for the analysis pertaining to the periods intended to be scrutinized). On the one hand, it has shown a significant decrease during the lockdown period in the population centers of Valladolid, Soria, Ponferrada, and Segovia; and on the other hand, Salamanca, León, Burgos, Zamora, and Palencia experienced relevant increases (Fig. 42).

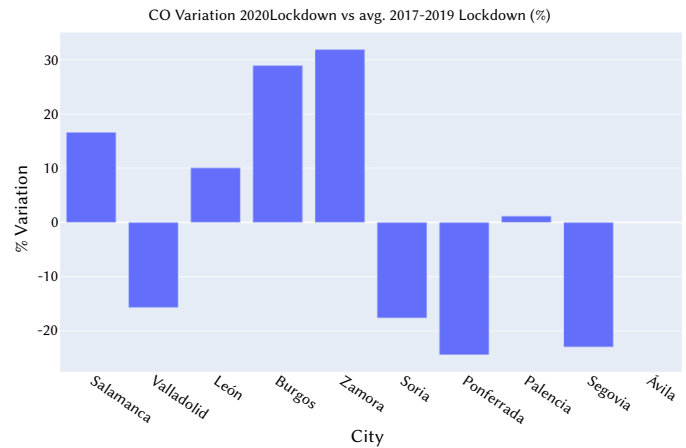


Fig. 42. Variation of CO between the 2020 lockdown period and the average of the same dates in 2017, 2018, and 2019 for the different population centers.

However, as seen in the heatmap in Fig. 43, Valladolid and Segovia recovered part of the lost values during the period following the lockdown, compared to the progress of the year so far. This is also due to the stationary trend, which causes the lowest levels to be reached during the period of the year in which the lockdown occurred (Fig. 33).

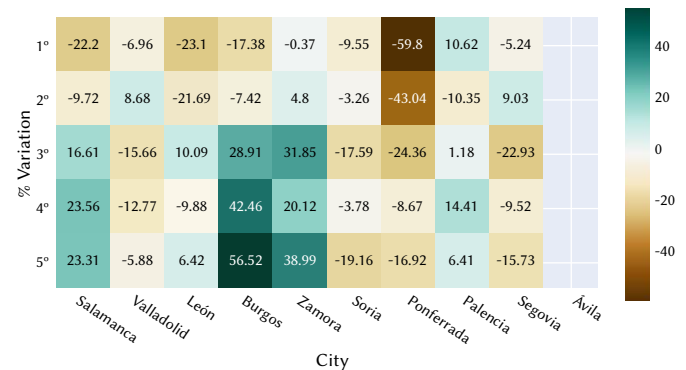


Fig. 43. Variation of CO in different time periods for the different population centers.

Delving into seasonality, it is seen that the variation between the lockdown period and the immediately preceding period, in general (except for Palencia), all population centers decrease their values, although in different ways depending on the impact during the lockdown.

Therefore, this leads to a decrease in 2020 compared to previous years in the population centers of Valladolid, Soria, Ponferrada, and Segovia; along with León, in the comparison of 2020 with 2019. The other group of provinces ended 2020 with a significant increase in their values compared to other years. This last point may be explained by the increasing trend of this pollutant over a few years, and one of the possible forecasts is that it will increase in the coming years, as can be seen in Fig. 33.

IV. CONCLUSIONS

The conclusions drawn from this study highlight the critical importance of having accurate predictions of pollutants, as this is essential for implementing measures to mitigate the damages caused by air pollution. Furthermore, it is important to investigate the causes, relationships, and trends of these pollutants in the short and long term, taking into account possible events that may alter their behavior, such as COVID-19. Accurate prediction allows for better information on air quality, enabling governmental organizations to prepare health plans that anticipate high levels of air pollution. Thus, they can adapt to any health event caused by atmospheric pollution phenomena.

The Prophet model has allowed us to make predictions that demonstrate a strong ability to forecast air quality in different spatial scenarios: various population centers with distinct regional characteristics; and temporal scenarios: in the short and long term, where attention should be paid to trends and seasonality. The possibilities for exploration with this approach are extensive in the field of air quality, surpassing ordinary prediction models such as LSTM or ARIMA. Particularly noteworthy are the cases of NO₂ y O₃ pollutants, where a high degree of accuracy is achieved, even for occasional peak levels. Moreover, they exhibit a precise prediction in any of the population centers according to the studied statistics. Furthermore, they show a precise forecast in any of the urban areas according to the statistics analyzed. This work therefore illustrates that Prophet has a broad capacity to forecast atmospheric pollution, and due to the fast training time and the lack of a complex system, it can be applied to other regions.

For the remaining pollutants discussed, several limitations regarding available data have hindered the model's ability to make accurate predictions, leading us to assess the importance of outliers, such as unanticipated meteorological events. Nonetheless, such intentionality was attributed to the constraints imposed by meteorological parameters on the adaptability of the model's prediction window.

In this study, emphasis is also placed on the analysis of trends for each pollutant and the seasonality they exhibit. This aids in achieving greater prediction accuracy and developing air quality plans that adapt accordingly.

Behavioral or restrictive events in society, such as COVID-19, disrupt the aforementioned factors, resulting in a significant impact on air quality and trends. In the current study, the implications of the lockdown due to the COVID-19 pandemic on air quality were assessed, in terms of variation and comparison among the studied population centers (the largest population centers in Castilla y León, Spain) during different periods surrounding the COVID-19 lockdown. It would be interesting to investigate, in future research, whether this event changed the behavior of the population and the interaction of pollutants with the environment.

The results showed a significant decrease in NO₂ y O₃ pollutants. This decline was not limited to the lockdown period, but the trend contributed to making 2020 one of the years with the lowest concentration of these pollutants in a long time. For the other pollutants, a decrease was also observed in most population centers, demonstrating how COVID-19 further emphasized the slope of the trend followed by these pollutants. It is worth noting that, in contrast to the other selected pollutants, CO experienced an increase in more than half of the studied population centers, confirming that its trend does not follow a decreasing pattern but rather stagnates. In general terms, with the exception of CO, a significant reduction in all atmospheric pollutants was observed during the closure period in the major population centers of Castilla y León. The findings of this study will be valuable for local municipal agencies and the administration of the Castilla y León region in order to establish rules and regulations aimed at enhancing and updating air quality in the future.

A. Limitations of the Study

The limitations of this study include geographical, as the study was limited to the provincial capitals and main cities of Castilla y León, and although the methodology of the experiment can be replicated, the results are only comparable and hardly extrapolable.

It should also be noted that the data are open data from government sources, where the accuracy of the sensors used to measure air quality is unknown. The study also does not include a review of meteorological data, such as wind gusts (speed and direction) or rainfall, which are relevant to pollutant dispersion.

V. FUTURE WORK LINES

Future research directions will focus on investigating the following aspects:

- To study and investigate the effects of wind gusts and their direction on the dispersion and concentration of these pollutants; identifying areas with higher pollutant concentrations would allow for the installation of green spaces in smart cities, which could improve air quality.
- To develop a federated learning architecture where different IoT devices for environmental monitoring can aggregate their readings and contribute to the training of models based on their location.
- To research Physics-Informed Neural Networks (PINNs) that are used to solve differential equations with applications in weather modeling, which may also help understand the movement of pollutant particles in the environment.
- To investigate long-term predictions based on the segmentation of time series into subseries that serve as input tokens to Transformer models and the independence of each channel. This approach would benefit from local information and long-term memory capabilities.

ACKNOWLEDGMENT

This work has been partially supported by the Institute for Business Competitiveness of Castilla y León, and the European Regional Development Fund under grant CCTT3/20/SA/0002 (AIR-SCity project).

REFERENCES

- [1] Bank, 2021. [Online]. Available: <https://data.worldbank.org/indicator/SP.RUR.TOTL.ZS>.
- [2] G. Duranton, D. Puga, "The growth of cities," *Handbook of economic growth*, vol. 2, pp. 781–853, 2014.
- [3] E. Commission, "Sustainable urban mobility in the eu: No substantial improvement is possible without member states' commitment," Jun 2020. [Online]. Available: <https://op.europa.eu/webpub/eca/special-reports/urban-mobility-6-2020/en/>.
- [4] W. H. Organization, 2016. [Online]. Available: https://www.euro.who.int/data/assets/pdf_file/0005/321971/Urban-green-spaces-and-health-review-evidence.pdf.
- [5] W. H. Organization, "Air pollution," 2023. [Online]. Available: <https://www.who.int/health-topics/air-pollution>.
- [6] Y. Zhu, J. Xie, F. Huang, L. Cao, "Association between short-term exposure to air pollution and covid-19 infection: Evidence from china," *Science of the total environment*, vol. 727, p. 138704, 2020.
- [7] P. Chamoso, A. González-Briones, S. Rodríguez, J. M. Corchado, "Tendencies of technologies and platforms in smart cities: a state-of-the-art review," *Wireless Communications and Mobile Computing*, vol. 2018, 2018.
- [8] T. Yigitcanlar, R. Mehmood, J. M. Corchado, "Green artificial intelligence: Towards an efficient, sustainable and equitable technology for smart cities and futures," *Sustainability*, vol. 13, no. 16, p. 8952, 2021.

- [9] V. Giannico, G. Spano, M. Elia, M. D'Este, G. Sanesi, R. Laforteza, "Green spaces, quality of life, and citizen perception in european cities," *Environmental Research*, vol. 196, p. 110922, 2021.
- [10] R. Casado-Vara, P. Novais, A. B. Gil, J. Prieto, J. M. Corchado, "Distributed continuous-time fault estimation control for multiple devices in iot networks," *IEEE Access*, vol. 7, pp. 11972–11984, 2019.
- [11] P. Pihkala, "Eco-anxiety and environmental education," *Sustainability*, vol. 12, no. 23, p. 10149, 2020.
- [12] R. Casado-Vara, A. Martín del Rey, R. S. Alonso, S. Trabelsi, J. M. Corchado, "A new stability criterion for iot systems in smart buildings: Temperature case study," *Mathematics*, vol. 8, no. 9, p. 1412, 2020.
- [13] T. Yigitcanlar, N. Kankanamge, M. Regona, A. Ruiz Maldonado, B. Rowan, A. Ryu, K. C. Desouza, J. M. Corchado, R. Mehmood, R. Y. M. Li, "Artificial intelligence technologies and related urban planning and development concepts: How are they perceived and utilized in australia?," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 6, no. 4, p. 187, 2020.
- [14] J. M. Corchado, F. Pinto-Santos, O. Aghmou, S. Trabelsi, "Intelligent development of smart cities: Deepint. net case studies," in *Sustainable Smart Cities and Territories*, 2022, pp. 211–225, Springer.
- [15] Y. Mezquita, A. González-Briones, R. Casado-Vara, P. Wolf, F. de la Prieta, A.-B. Gil-González, "Review of privacy preservation with blockchain technology in the context of smart cities," in *Sustainable Smart Cities and Territories*, 2022, pp. 68–77, Springer.
- [16] Y.-S. Chang, H.-T. Chiao, S. Abimannan, Y.-P. Huang, Y.-T. Tsai, K.-M. Lin, "An lstm-based aggregated model for air pollution forecasting," *Atmospheric Pollution Research*, vol. 11, no. 8, pp. 1451–1463, 2020.
- [17] J. M. Corchado, "Technologies for sustainable consumption - researchgate.net," Apr 2021.
- [18] R. López-Blanco, R. S. Alonso, J. Prieto, S. Trabelsi, "Automating the implementation of unsupervised machine learning processes in smart cities scenarios," in *Distributed Computing and Artificial Intelligence, Special Sessions, 19th International Conference*, 2023, pp. 71–80, Springer.
- [19] T. Dimri, S. Ahmad, M. Sharif, "Time series analysis of climate variables using seasonal arima approach," *Journal of Earth System Science*, vol. 129, no. 1, pp. 1–16, 2020.
- [20] R. López-Blanco, J. H. Martín, R. S. Alonso, J. Prieto, "Time series forecasting for improving quality of life and ecosystem services in smart cities," in *Ambient Intelligence—Software and Applications—13th International Symposium on Ambient Intelligence*, 2023, pp. 74–85, Springer.
- [21] A. Hasnain, Y. Sheng, M. Z. Hashmi, U. A. Bhatti, A. Hussain, M. Hameed, S. Marjan, S. U. Bazai, M. A. Hossain, M. Sahabuddin, *et al.*, "Time series analysis and forecasting of air pollutants based on prophet forecasting model in jiangsu province, china," *Frontiers in Environmental Science*, p. 1044, 2022.
- [22] J. Shen, D. Valagolam, S. McCalla, "Prophet forecasting model: A machine learning approach to predict the concentration of air pollutants (pm2. 5, pm10, o3, no2, so2, co) in seoul, south korea," *PeerJ*, vol. 8, p. e9961, 2020.
- [23] G. Swamy, S. Nagendra, U. Schlink, "Impact of urban heat island on meteorology and air quality at microenvironments," *Journal of the Air & Waste Management Association*, vol. 70, no. 9, pp. 876–891, 2020.
- [24] J. Ngarambe, S. J. Joen, C.-H. Han, G. Y. Yun, "Exploring the relationship between particulate matter, co, so2, no2, o3 and urban heat island in seoul, korea," *Journal of Hazardous Materials*, vol. 403, p. 123615, 2021.
- [25] G. Miskell, W. Pattinson, L. Weissert, D. Williams, "Forecasting short-term peak concentrations from a network of air quality instruments measuring pm2. 5 using boosted gradient machine models," *Journal of environmental management*, vol. 242, pp. 56–64, 2019.
- [26] W.-W. Li, R. Orquiz, J. H. Garcia, T. T. Espino, N. E. Pingitore, J. Gardea-Torresdey, J. Chow, J. G. Watson, "Analysis of temporal and spatial dichotomous pm air samples in the el paso-cd. juarez air quality basin," *Journal of the Air & Waste Management Association*, vol. 51, no. 11, pp. 1551–1560, 2001.
- [27] S. Fei, R. A. Wagan, A. Hasnain, A. Hussain, U. A. Bhatti, E. Elahi, "Spatiotemporal impact of the covid-19 pandemic lockdown on air quality pattern in nanjing, china," *Frontiers in Environmental Science*, p. 1548, 2022.
- [28] Y. Wang, S. Zhu, C. Li, "Research on multistep time series prediction based on lstm," 10 2019, pp. 1155–1159.
- [29] M. C. Turner, Z. J. Andersen, A. Baccarelli, W. R. Diver, S. M. Gapstur, C. A. Pope III, D. Prada, J. Samet, G. Thurston, A. Cohen, "Outdoor air pollution and cancer: An overview of the current evidence and public health recommendations," *CA: a cancer journal for clinicians*, vol. 70, no. 6, pp. 460–479, 2020.
- [30] "Datos abiertos de castilla y león. calidad del aire (por días)." https://datosabiertos.jcyl.es/web/jcyl/set/es/medio-ambiente/calidad_aire_historico/1284212629698. Accessed: 2023-07-07.
- [31] W. Robson, "The math of prophet," Jun 2019. [Online]. Available: <https://medium.com/future-vision/the-math-of-prophet-46864fa9c55a>.
- [32] S. J. Taylor, B. Letham, "Forecasting at scale," 09 2017, doi: 10.7287/peerj.preprints.3190v2.
- [33] AEMET, "Opendata aemet." [Online]. Available: <https://opendata.aemet.es>.
- [34] S. Abdullah, A. A. Mansor, N. N. L. M. Napi, W. N. W. Mansor, A. N. Ahmed, M. Ismail, Z. T. A. Ramly, "Air quality status during 2020 malaysia movement control order (mco) due to 2019 novel coronavirus (2019-ncov) pandemic," *Science of The Total Environment*, vol. 729, p. 139022, 2020, doi: <https://doi.org/10.1016/j.scitotenv.2020.139022>.
- [35] R. Bao, A. Zhang, "Does lockdown reduce air pollution? evidence from 44 cities in northern china," *Science of The Total Environment*, vol. 731, p. 139052, 2020, doi: <https://doi.org/10.1016/j.scitotenv.2020.139052>.
- [36] E. Bontempi, C. Carnevale, A. Cornelio, M. Volta, A. Zanoletti, "Analysis of the lockdown effects due to the covid-19 on air pollution in brescia (lombardy)," *Environmental Research*, vol. 212, p. 113193, 2022, doi: <https://doi.org/10.1016/j.envres.2022.113193>.



Raúl López Blanco

Predocctoral researcher, MSc in Web Site Management and Engineering (2021) by the International University of La Rioja (UNIR) (Spain) and BSc in Computer Engineering (2019) by the University of Salamanca (Spain). Researcher at the AIR Institute and member of the BISITE research group since 2019. Teacher in Internet of Things MSc and Industry 4.0 MSc at the University of La Rioja (UNIR)

since 2022. Software developer with experience in programming languages such as C++, Java, C# JavaScript, web development, blockchain and databases, among others. His main research topics focus on traceability systems, databases, IoT and distributed systems. Raúl has published papers on IoT and AI and has participated in more than 10 research projects related to Smart Data, Smart Farming, Smart Healthcare, Deep Learning, Traceability, IoT and IIoT.



Miguel Chaveinte García

Last year student of the Degree in Computer Engineering at the University of Valladolid (UVA). Currently, he combines his studies with his work at Air Institute. His work within the organization focuses on several areas ranging from the development of backend systems to the research of new Deep Learning algorithms and explainable Artificial Intelligence, along with data analysis processes.



Ricardo S. Alonso, PhD.

Ricardo holds a PhD in Computer Engineering from the University of Salamanca (Spain), a MSc in Intelligent Systems from the same university and a Telecommunications Engineering degree from the University of Valladolid (Spain). He also holds a Master's degree in Business Administration (MBA) from the International University of La Rioja (Spain). On the one hand, he is currently a researcher at the AIR Institute and he is part of the BISITE Research Group (Bioinformatics, Intelligent Systems and Educational Technology) at the University of Salamanca. He has published more than 15 research articles in international journals and more than 35 in international conferences. During this career, he has collaborated in more than 40 regional, national, and European R&D projects, having been the Principal Investigator in regional and European projects funded by the Horizon 2020 program. His research interests include Internet of Things, Edge Computing, Distributed Ledger and Blockchain technologies, Embedded Systems, Indoor Location Systems, Cloud Computing and Artificial Intelligence. On the other hand, he has worked for more than 10 years in private enterprise as an entrepreneur, technologist, and researcher in the field of Wireless Sensor Networks and Real Time Location Systems, acting as Deputy R&D Director, R&D Director, Sales Manager and Co-Founder in different companies. He is also a lecturer at the International University of La Rioja, where he is Academic Director of the Interuniversity Master's Degree in AI for the Energy and Infrastructure Sector.



Javier Prieto, PhD

Full Professor at the University of Salamanca in the Department of Computer Science and Automation. He is a member of BISITE and ATA (University Institute for Research in Art and Technology of Animation). BSc in Telecommunications Engineering and PhD in Information and Communication Technologies (University of Valladolid). Since 2007, Javier Prieto has worked in diverse public and private research centres: Fundación Centro para el Desarrollo de las Telecomunicaciones en Castilla y León (CEDETEL), University of Valladolid (UVa), or the Massachusetts Institute of Technology (MIT) in Cambridge, MA (USA). He has received competitive grants as Postdoctoral Researcher by the University of Salamanca, or Postdoctoral Researcher Torres Quevedo by the Ministry of Science and Innovation. He has published over 150 articles in international journals, books, and congresses, and is the author of 2 national patents. He is the editor in chief of the Smart Cities journal, at its IoT Section, and a senior editor of the journals IEEE Communications Letters and Wireless Communications and Mobile Computing, a guest editor of numerous special issues in different journals, and a member of the scientific committee of the Advances in Distributed Computing and AI Journal. He is co-founder and Technical Secretariat at AIR Institute, IEEE Senior Member, board member of IEEE Spain Section, Head of the IEEE Blockchain Initiative, Operations Chair of the IEEE GLOBECOM 2021 congress, member of the TC Cognitive Networks (TCCN) Technical Committees of the IEEE ComSoc and RFID Technologies Committee (MTT-24) of the IEEE MTT-S. He has been president of the technical program committee of the IEEE ICUWB 2017. Teaching activity: Full Professor at the University of Salamanca and Associate Professor at the University of Valladolid. Javier has been professor of the Master in Big Data Science at the University of Valladolid and the Master in Industry 4.0 (IIoT) at the International University of La Rioja. Research interests: social computing and AI for smart cities, location and navigation technologies in both indoor and outdoor environments, and Bayesian inference techniques for improving social welfare and sustainable development. He has coordinated and participated in over 100 national and international R&D projects.



Juan M. Corchado, PhD.

PhD in Computer Sciences (University of Salamanca). PhD in Artificial Intelligence (University of the West of Scotland). Director of the BISITE Research Group at the University of Salamanca. Director of the IoT Digital Innovation Hub. Visiting Professor at the Osaka Institute of Technology. Visiting Professor at the University Malaysia Kelantan. Member of the Advisory group on Online Terrorist Propaganda (European Counter Terrorism Centre, EUROPOL). He is currently University Professor, Director of Postgraduate Studies in Security, Information Systems, social media, Mobile Telephony, Digital Animation, IoT and Blockchain, having directed 17 doctoral theses. He is also the director of the IoT Digital Innovation Hub and Academic Director of the University Institute for Research in Animation Art and Technology at the University of Salamanca. He is a Visiting Professor at the Osaka Institute of Technology (Japan). Previously he has been Vice-rector of Research and Transfer at the University of Salamanca (2013-2017), Director of the University of Salamanca Science Park (2013-2017), Deputy Director of the School of Computer Engineering at the University of Vigo (1999-00), researcher at the University of Paisley (1995-98) and Dean in the Faculty of Sciences at the University of Salamanca (2008-2013). He worked as a programmer at the Oceanographic Laboratory in Plymouth (United Kingdom) between 1993 and 1998. Juan M. Corchado is President of the Spanish chapter of the association IEEE Systems, Man and Cybernetics and member of the Advisory Group on Online Terrorist Propaganda of EUROPOL. He is also editor-in-chief of ADCAIJ (Advances in Distributed Computing and Artificial Intelligence Journal) and IJDA (International Journal of Digital Contents and Applications). He is co-author of more than 350 books, book chapters, articles in international journals, etc., most of them presenting both practical and theoretical aspects related to Hybrid Artificial Intelligence Systems, Distributed Computing, Biomedicine, Ambient Intelligence, Wireless Network Systems, Multiagent Systems, etc. He has worked on more than 100 research projects (European, National and Autonomous) and research contracts (article 83). He has been the coordinator and Principal Investigator in different H2020 and Interreg PocTep project, as well as numerous national projects.