International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V71I10P201
Volume 71 Issue 10, 1-10, October 2023


*Original Article*

# Application of the Use of Time Series Models: Tropospheric Nitrogen Dioxide ($NO_2$) in Different Meteorological Systems in Two Districts of the City of Lima

Airton Fabrizio Molina-Cueva[1], Renzo Aaron Cueva-Roldan[1], Yvan Jesus Garcia-Lopez[1], Juan Carlos Quiroz-Flores[1*]

[1]*Faculty of Engineering, Career of Industrial Engineering, University of Lima, Perú.*

[*]*Corresponding Author : jcquiroz@ulima.edu.pe*

Received: 02 July 2023 Revised: 06 September 2023 Accepted: 07 September 2023 Published: 03 October 2023

***Abstract*** - *This research will address air pollution, a severe problem in all world cities, because it negatively affects people's health and deteriorates the ecosystem. $NO_2$ is a gas linked to acid rain formation and various reactions with greenhouse gases. Meteorological variables influence the behavior of tropospheric $NO_2$ concentration. During the period of confinement due to the COVID-19 pandemic, the concentration levels of pollutants dropped abruptly, which meant relief for the ecosystem. The application of Time Series models allows us to graphically identify the concentration of contaminants in various areas and make accurate forecasts to mitigate environmental problems in the future. The research analysis shows that the SARIMA model effectively forecasts the pollutant concentration in the San Borja and San Martin de Porres districts in Lima. Error tests such as $R^2$, MAE, MAPE, MSE, and RSME, as well as Dickey-Fuller Test, AIC, BIC, Skew, and Kurtosis, provide information on the performance of the SARIMA model and show that it is the most suitable.*

***Keywords*** - *Air pollution, Time series, ARIMA, SARIMA, and Tropospheric $NO_2$.*

## 1. Introduction

Air pollution is a problem of growing concern that affects all cities in the world. This problem cannot be avoided due to the growth of industries, population increase, and technological advances [1]. Also, the transport sector is the main contributor to air pollution, highlighting that polluted air has mild and severe effects on all living beings depending on the time they are exposed to, the concentration of pollutants, and their health [2]. An analysis in Lima, specifically in the San Martin de Porres and San Borja districts, shows that industrial activities and the automotive fleet are the leading causes of air pollution. In these districts, they found six types of pollutants in the air: CO, $SO_2$, $O_3$, $NO_2$, PM10, and PM2.5. Air pollution impacts human health and quality of life and can cause cardiovascular diseases, allergies, and respiratory problems [3]. $NO_2$ is a polluting gas generated by anthropogenic production and the environment that harms humans and their ecosystem. Even in low concentrations, it can cause inflammatory and respiratory diseases in the respiratory system, harm vegetation development, and contribute to the generation of acid rain, mist, and smog [4]. Meteorological variables differ by origin and trajectory (air temperature, relative humidity, precipitation, wind speed, and wind direction). They can alter the behavior of the composition and concentration of pollutants. According to $NO_2$ research, at high temperatures, the NO molecules emitted from the exhaust pipe of vehicles oxidize in the atmosphere, forming $NO_2$ [5]. An analysis of the previous and current levels of pollutants, such as $NO_2$, helps predict the contaminant's behavior trend. Various studies explain that the behavior of pollutants has a pattern of dynamic variability, making them good candidates for calculating forecasts of their concentrations by applying time series models. Time series models provide considerable advantages for short and long-term forecasting [6]. During the COVID-19 pandemic, the concentration levels of the pollutant decreased drastically, reaffirming anthropogenic factors as the main cause of the altered behavior of $NO_2$ concentration [7].

Adequate forecasting of the pollutant would make it possible to identify the reasons for the increase in concentration in each area and establish contingency plans to mitigate its exposure. In Peru, the agencies responsible for the health of people and the protection of the environment have established the maximum annual $NO_2$ concentration limit (ECA) of 100ug/m$^3$ as "healthy" [8]. In the districts of the city of Lima, it has become evident that the annual $NO_2$ concentration levels diverge questionably among them. While some districts may have a stable annual average value, others may reach twice as high in the same period. The districts of San Borja and San Martin de Porres are candidates for study, possessing remarkable qualities such as level of development and size of territory, respectively. This research seeks to present an approach based on time series methods to calculate a forecast of $NO_2$ concentration in the districts of San Borja and San Martin de Porres and to determine the most appropriate model. Currently, there are few studies on using time series models for $NO_2$ concentration forecasting in this country, so this research highlights a crucial role in the environmental field that will complement and challenge traditional methods. In addition, it seeks to get the interest of district governments and municipalities to implement environmental policies for the benefit of the people.

## 2. Literature Review

Several previous studies that are relevant to this study have been reviewed in this section.

In the context of the COVID-19 pandemic, one study used the number of patients with disease infection cases as data. They found that they evolved with a dynamic fluctuation trend due to situations such as epidemic prevention and control. For this type of data, the models best suited to forecast with high accuracy are the ARIMA and SARIMA methods, which, despite not needing much mathematics or statistics, can capture the regularity of the data, such as the short-term changing trends in time series [6].

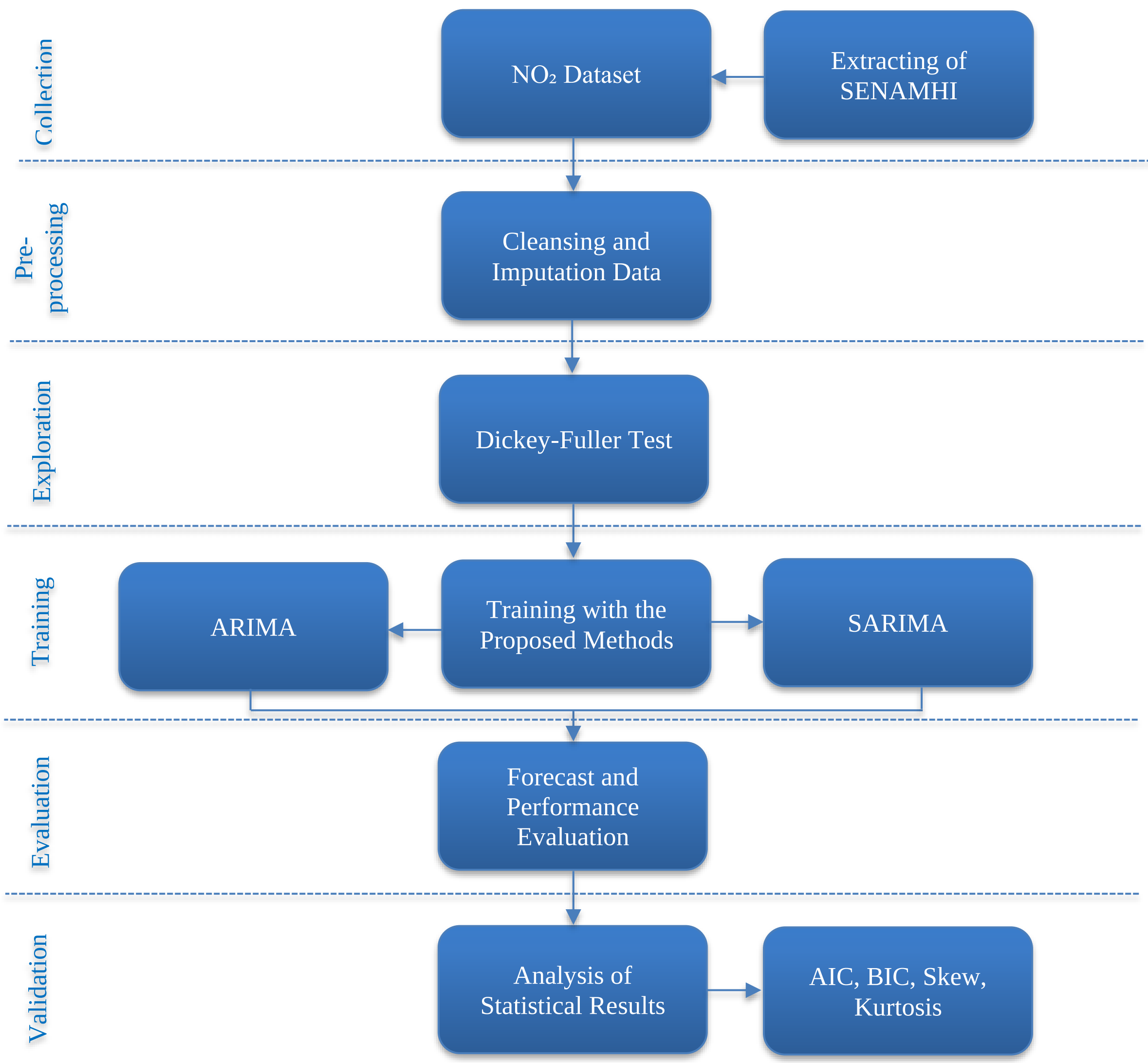


**Fig. 1 Proposed methodology of forecast models**

Another research whose main objective is to find the best model for $CO_2$ prediction in Peru by comparing the ARIMA and Artificial Neural Networks models; so, to determine which of these two models is the best, he used forecast errors as RMSE, MAE, and MAPE [9]. Using these indicators, it was possible to verify that the Artificial Neural Network model is the best model that adapts to predicting $CO_2$ emissions in Peru.

By applying Machine Learning models, a group of researchers conducted a workshop to predict the air quality index in Delhi and National Capital Region cities. They measured their results based on MAPE, MAE, RMSE, and MSE values [10].

A study conducted in India's city mentions that the ARIMA model can predict the values of air pollutants based on historical values. It also says that it is an effective model when it is required to know if the pollutant values exceed the limits established by the WHO [11].

On the other hand, in research on electric power, a comparison was made between four-time series models to determine which gave the best electricity generation forecast in grid-connected PV systems. The models used were SARIMA, SARIMAX, modified SARIMA, and Artificial Neural Network. According to the results of this research, it was determined that the SARIMAX, modified SARIMA, and Artificial Neural Network models perform better than the SARIMA model.

The importance of using external factors such as solar radiation for time series models and using forecast errors such as the root mean square error (RMSE) to evaluate the performance of each model is highlighted [12].

In addition, another study developed in their research work a dual SARIMA model that can improve the accuracy in forecasting short-term load for electric power planning in Malaysia. A comparison was made with the single SARIMA model, showing that the double SARIMA model is more accurate than the single SARIMA model. This is evidenced by the MAPE, as the MAPE inside and outside the sample is reduced by 10.23% and 57%, respectively [13].

Other authors show us in their research work the use of analytical models such as SARIMA and Prophet to create a system that can be able to forecast the pollution levels of pollutants such as RSPM, $NO_2$, $SO_2$, and SPM in a considerable confidence interval in the city of Bhubaneswar, they made a comparison using RMSE and MSE indicators and concluded that both SARIMA and Prophet models are very accurate when forecasting; However, the best model for this work is their proposed logarithmic transformation Prophet model, which obtained the lowest values for the RMSE and MSE indicators [14].

## 3. Materials and Methods

### *3.1. Data Collection*

The areas where this study will be carried out are characterized by high vehicular activity and movement of people. The air quality monitoring station is in the Ecological Park of the district of San Martin de Porres, considered one of the most polluted districts of Metropolitan Lima, and is located 500m from Universitaria Avenue, known for being the longest road in Lima and having high traffic and stores in its vicinity. In the case of the station located at the Sports Center Limatambo in the district of San Borja, it is located in a "green" zone comprised of commercial and residential areas in the surrounding area. The primary source of information is the reports made by SENAMHI through the Automatic Air Quality Monitoring Network (AAQMN) located in these districts, where one of the most important atmospheric pollutants will be the study variable and the meteorological variables. The data includes meteorological variables and the concentration of the pollutant $NO_2$. The meteorological variables are comprised of Air Temperature, Relative Humidity, Precipitation, Wind Speed, and Wind Direction. The dataset includes hourly data for four years and seven months between 2018 and 2022. The null data are cleaned from the original dataset.

### *3.2. Data Process*

Due to the high amount of missing data (null, empty, or unavailable), a manual imputation with substitute values was performed. The hourly imputation method consisted of averaging the data available during the 24 hours of the day and placing them in the missing hourly positions.

### *3.3. Python Language*

Python is a freely available, dynamic, and interactive programming language for creating and prototyping large applications [15]. Its advantages of use are based on its power to run multiple Windows applications, web pages, network servers, etc. [16]. Its popularity is mainly given by its facility to perform Big Data, Machine Learning, and Data Science. It relies on the excellent support of libraries that allow coding without the need to implement algorithms, reducing development time and execution speed [15,16].

### *3.4. Time Series Models*

A time series is defined as the values of a variable characterized by having an ordered sequence with equally spaced time intervals between them [17]. The behavior of the data over time does not always have a defined shape and may present trend and seasonal variations [18]. Time series models follow a predictive approach by analyzing and processing past observations of the variable of interest to obtain forecasts appropriate to the time series [17, 18]. The structural time series model considers random residuals as the natural variability in the trend of the series to perform the autoregressive moving average process [19].

**Table 1. Tropospheric $NO_2$ pollutant and meteorological variables**

| Item | Variable | Definition |
|---|---|---|
| Primary | $NO_2$ | It is an atmospheric toxic gas that can come from road traffic, industries, forest fires, and volcanic eruptions. |
| Secondary | Air temperature | It measures the amount of heat in the air. |
| | Relative humidity | The amount of water vapor in the air is based on the maximum amount that can be contained at a given temperature. |
| | Precipitation | It is liquid or solid water formed in the atmosphere that returns to the earth's surface in the form of hydrometeor. |
| | Wind speed | It is the distance an air particle travels in a given time. |
| | Wind direction | It is the direction in which the wind blows. |

*3.4.1. ARIMA Model*

ARIMA (Autoregressive Integrated Moving Average) is a traditional algorithm for analyzing non-stationary time series. Unlike other regression models, ARIMA uses past values and stochastic error terms to explain the time series. These models, known as ARIMA models, combine Autoregressive (AR) and Integration (I), which refer to the inverse process of differencing to produce the prediction and Moving Average (MA) components [20]. The ARIMA model is denoted ARIMA (p,d,q) where "p" represents the number of observations of the autoregressive process, "d" is the order of data stationarization, and "q" is the order of the moving average method [22]. The general form of the ARIMA (p,d,q) model is described as follows:

$$\begin{aligned} y_t = {} & \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} \\ & + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \end{aligned} \tag{1}$$

*3.4.2. SARIMA Model*

It is an extension of the ARIMA model that differs in its ability to model time series with seasonal patterns. The SARIMA (Seasonal Autoregressive Integrated Moving Average) model considers the seasonal characteristic of the data using additional components for the stationarity adjustment of the time series. It is based on applying mathematical models to non-stationary time series after data smoothing, which estimates and extrapolates by analyzing past and present historical patterns to forecast the future [22]. The SARIMA model is denoted similarly to ARIMA but with an additional seasonal term $(P,D,Q)^m$ representing the numbers for handling stationarity and periodicity of stationarity [23]. The general formula of the SARIMA model $(p,d,q)(P,D,Q)^m$ is described as follows:

$$\begin{aligned} y_t = {} & \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} \\ & + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \\ & + \Phi_1 y_{t-m} - \Phi_2 y_{t-2m} - \cdots - \Phi_p y_{t-pm} \\ & - \Theta_1 e_{t-m} - \Theta_2 e_{t-2m} - \cdots - \Theta_q e_{t-qm} \end{aligned} \tag{2}$$

### *3.5. Tools and Model Configuration*

The data and the SARIMA forecasting model were programmed in Python using the Pandas, NumPy, Matplotlib, Seaborn, Statsmodels, Scikit-learn, and Pmdarima libraries. Python programming allows you to handle, clean, and process large amounts of data. The libraries allow statistical analysis of the data set to obtain the coefficients $R^2$, MAE, MSE, RMSE, MAPE, Dickey-Fuller Test, and the Diagnostic Model in addition to the forecast. While the $R^2$ coefficient measures the proportion of the variability of the dependent variable, the MAE, MSE, and RMSE evaluate the discrepancy between the values predicted by the algorithm and the actual values of the data, and the MAPE, the difference expressed as a percentage. The statistical indicators AIC, BIC, Skew, and Kurtosis were obtained with the prognosis. Low AIC and BIC values represent that the model is adequate; Skew shows the trend of the skewness of the data distribution, and Kurtosis is the concentration level of the data around the mean.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2}{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2} \tag{3}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y}_i) \tag{4}$$

$$MSE = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2}{n} \tag{5}$$

$$RSME = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2}{n}} \tag{6}$$

$$MAPE = \frac{100}{n} \times \sum_{i=1}^{n} \left| \frac{(Y_i - \bar{Y}_i)}{Y_i} \right| \tag{7}$$

$$AIC = -2logL(\theta) + 2K \tag{8}$$

$$BIC = -2logL(\theta) + Klogn \tag{9}$$

$$Skew = \frac{1}{n}\sum_{i=1}^{n}\frac{(X_i - \bar{X})^3}{s^3} \quad (10)$$

$$Kurtosis = \frac{1}{n}\sum_{i=1}^{n}\frac{(X_i - \bar{X})^4}{s^4} \quad (11)$$

Where n is the number of observations, in equations (3-7), the variable Yi is the current value, Ȳi represents the model prediction value, and Ŷ refers to the mean of the observed values. In equations (8) and (9), L(θ) refers to the log-likelihood, and K is the number of model parameters. In equations (10) and (11), the variables Xi, X̄i, and s represent the sample's mean and standard deviation values, respectively.

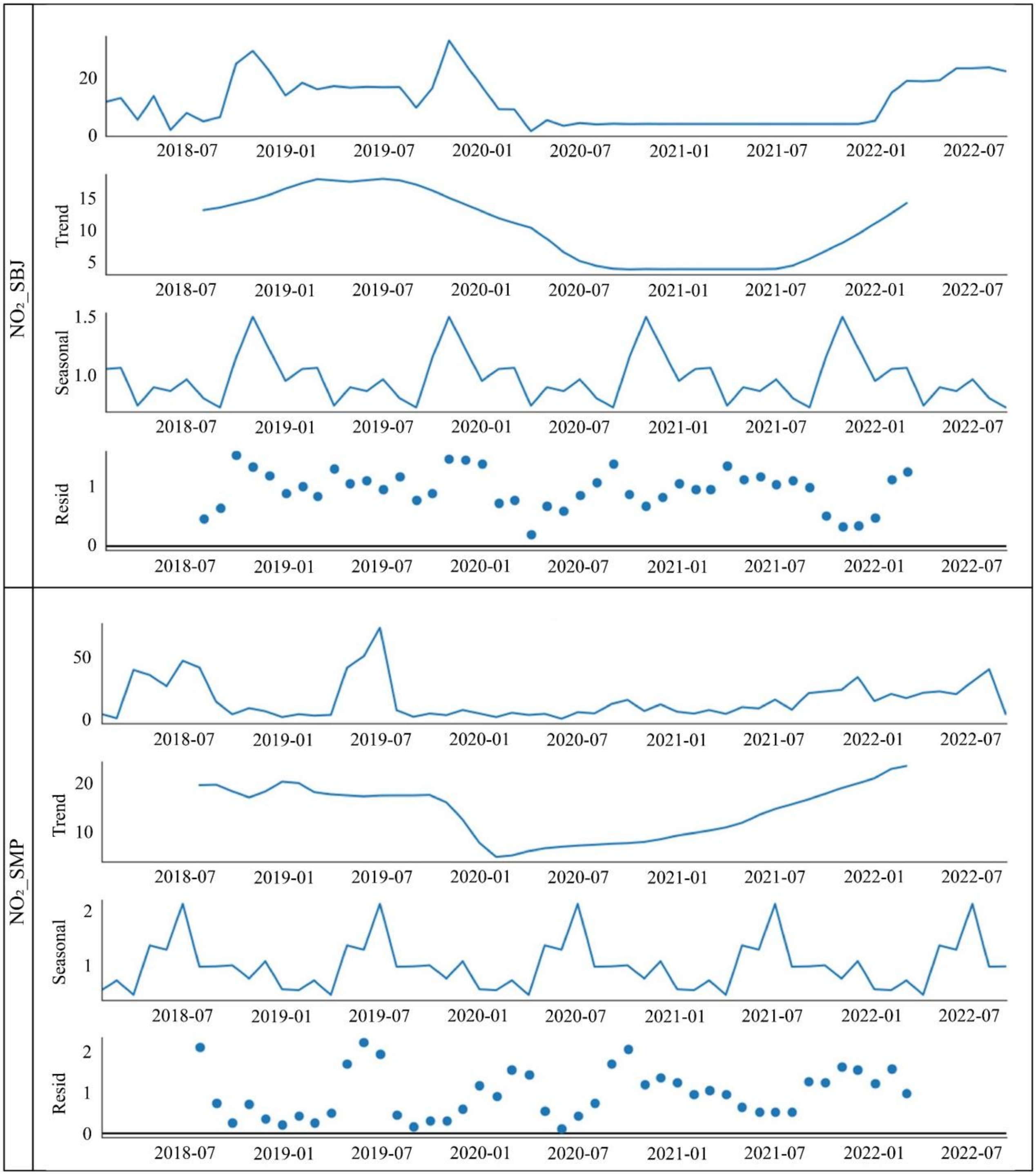


**Fig. 2 Properties of the time series for both districts**

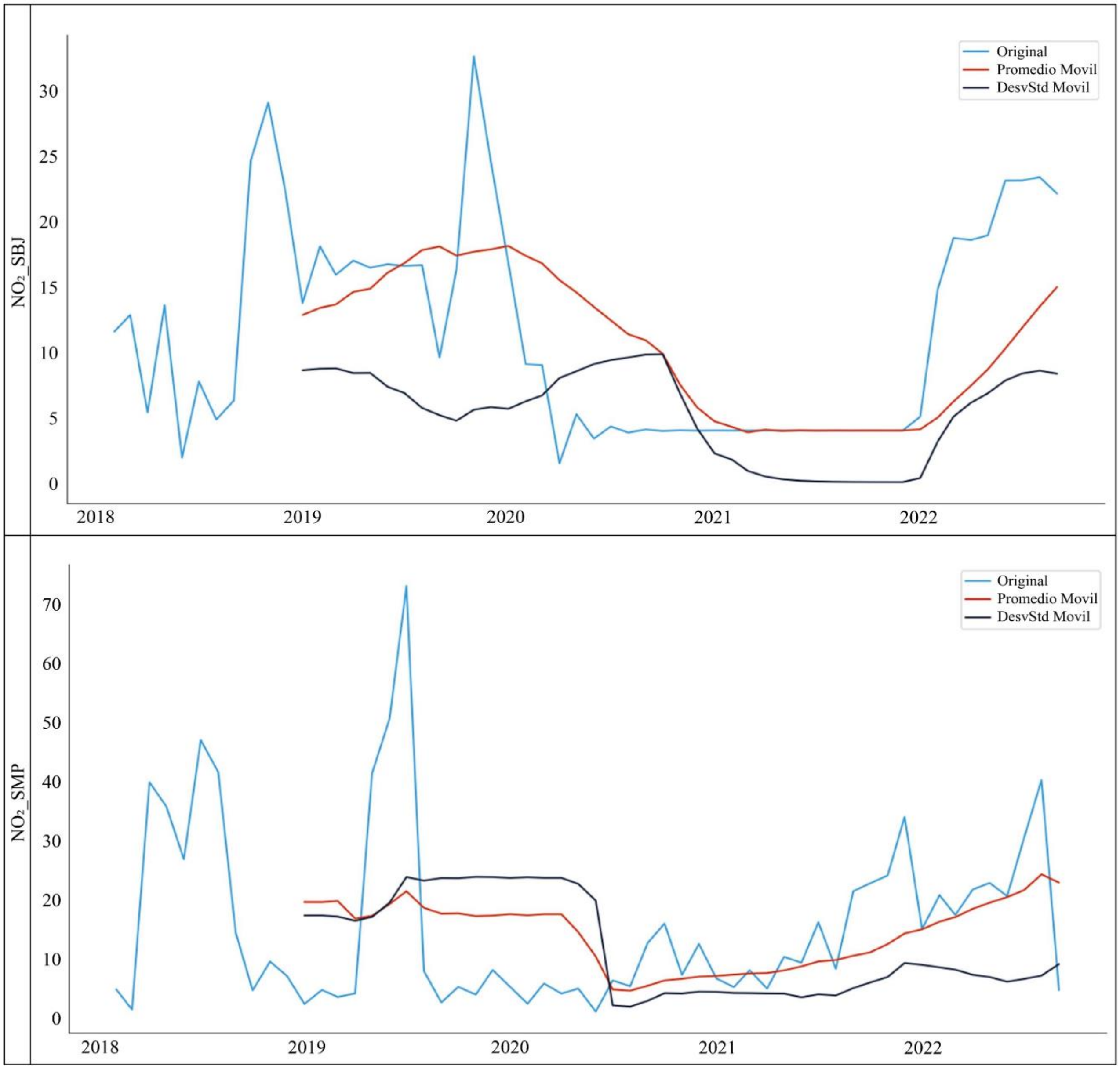


**Fig. 3 Moving average and standard deviation of $NO_2$ in both districts**

## 4. Results and Discussion

This section will present $NO_2$ forecasts for San Borja and San Martin de Porres districts using data collected from meteorological stations from 2018 to 2022. These data are processed based on model requirements. The imputation method was used for missing values by averaging the previous month with the following month.

In addition, before modeling the data, a visualization of the time series had to be performed to see if they have seasonality; this visualization facilitates the choice of parameters for the models. Likewise, the data for training and testing were divided into 60% and 30%, respectively. This is done to understand the pattern of behavior of the time series in both districts. Finally, statistical regressors were imported to measure the statistical errors for each district.

### *4.1. Seasonality Test*

In the present investigation, two methods were applied to each time series to determine whether they were stationary. The first method determines the moving statistics of the mean and standard deviation, while the second applies the augmented Dickey-Fuller Test in Table 2. The results of the moving statistics for the district of San Borja showed that during the year 2021, the time series had a negative trend in its moving average and standard deviation. Still, entering the year 2022, it is observed that both achieved growth.

**Table 2. Dickey-Fuller test**

| Dickey-Fuller's Test | SBJ | SMP |
|---|---|---|
| Test Statistic | -2,345,450 | -4,007,826 |
| p-value | 0.157728 | 0.001370 |
| #Lags Used | 0.000000 | 0.000000 |
| Number of Observations Used | 55,000,000 | 55,000,000 |
| Critical Value (1%) | -3,555,273 | -3,555,273 |
| Critical Value (5%) | -2,915,731 | -2,915,731 |
| Critical Value (10%) | -2,595,670 | -2,595,670 |

Likewise, for the district of San Martin de Porres, the negative trend appears in mid-2020 for both the moving average and the standard deviation; however, by mid-2020, there is a constant growth in the moving average and the standard deviation.

On the other hand, the Augmented Dickey-Fuller Test was also applied to the data of both districts, establishing a significance level before the test and drawing conclusions based on the resulting p-value. The results obtained in each time series indicate that in the case of the district of San Borja, there is a p-value=0.157728, which means that it is very likely that the data is not stationary. Meanwhile, for the district of San Martin de Porres, a p-value=0.001370 was obtained, indicating that the data for this district is stationary.

### *4.2. Identification of the Best Model*

In this part, the models will be evaluated using the Akaike (AIC) and Bayesian Information Criterion (BIC) for each district's ARIMA and SARIMA models. These criteria are used to determine which model is the best by comparing their AIC and BIC so that the model with the lowest AIC and BIC will be the best model for their pollutant forecast. Based on the results in Table 3 and Table 4, the best model that will be used for $NO_2$ forecasting is SARIMA. Regarding the Skew for the SARIMA model, in both districts, negative values are presented; therefore, they show negative asymmetry curves. It also represents that the values of the mean of the concentration of these districts are lower than their medians, and comparing their Kurtosis for the model SARIMA, the district that presents a better result is San Borja, which means that it has fewer outliers compared to the values of the district of San Martin de Porres.

**Table 3. MAPE, MAE, MSE, and RMSE values for both districts**

| Statistical Error | SBJ | SMP |
|---|---|---|
| MAPE | 0.44 | 0.91 |
| MAE | 3.32 | 8.61 |
| MSE | 30.06 | 210.15 |
| RMSE | 5.48 | 14.49 |

**Table 4. AIC and BIC values for both districts**

| | ARIMA | | SARIMA | |
|---|---|---|---|---|
| District | AIC | BIC | AIC | BIC |
| SBJ | 341.761 | 343.769 | 279,767 | 283,289 |
| SMP | 453,096 | 455,103 | 368,038 | 375,082 |

**Table 5. Skew and Kurtosis for SARIMA model**

| Statistical Metrics | SBJ | SMP |
|---|---|---|
| Skew | -0.24 | -0.73 |
| Kurtosis | 2.70 | 3.45 |

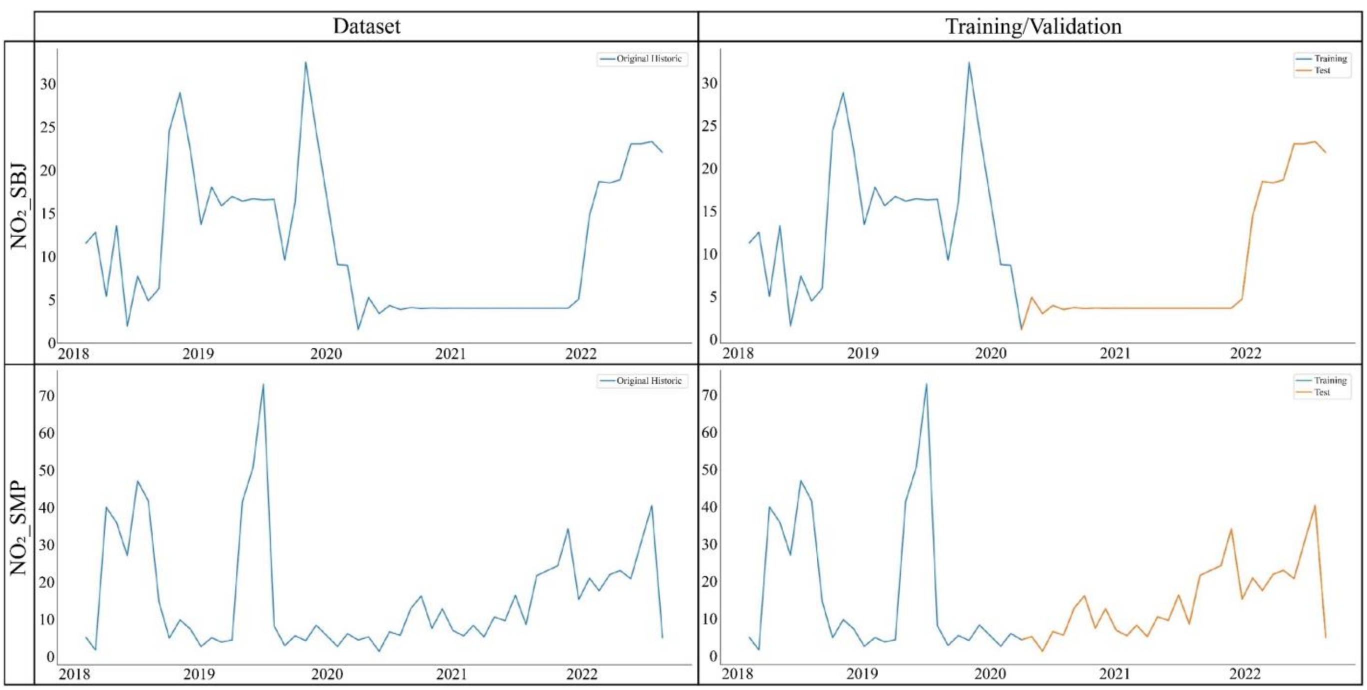


**Fig. 4 The dataset to be trained and validated**

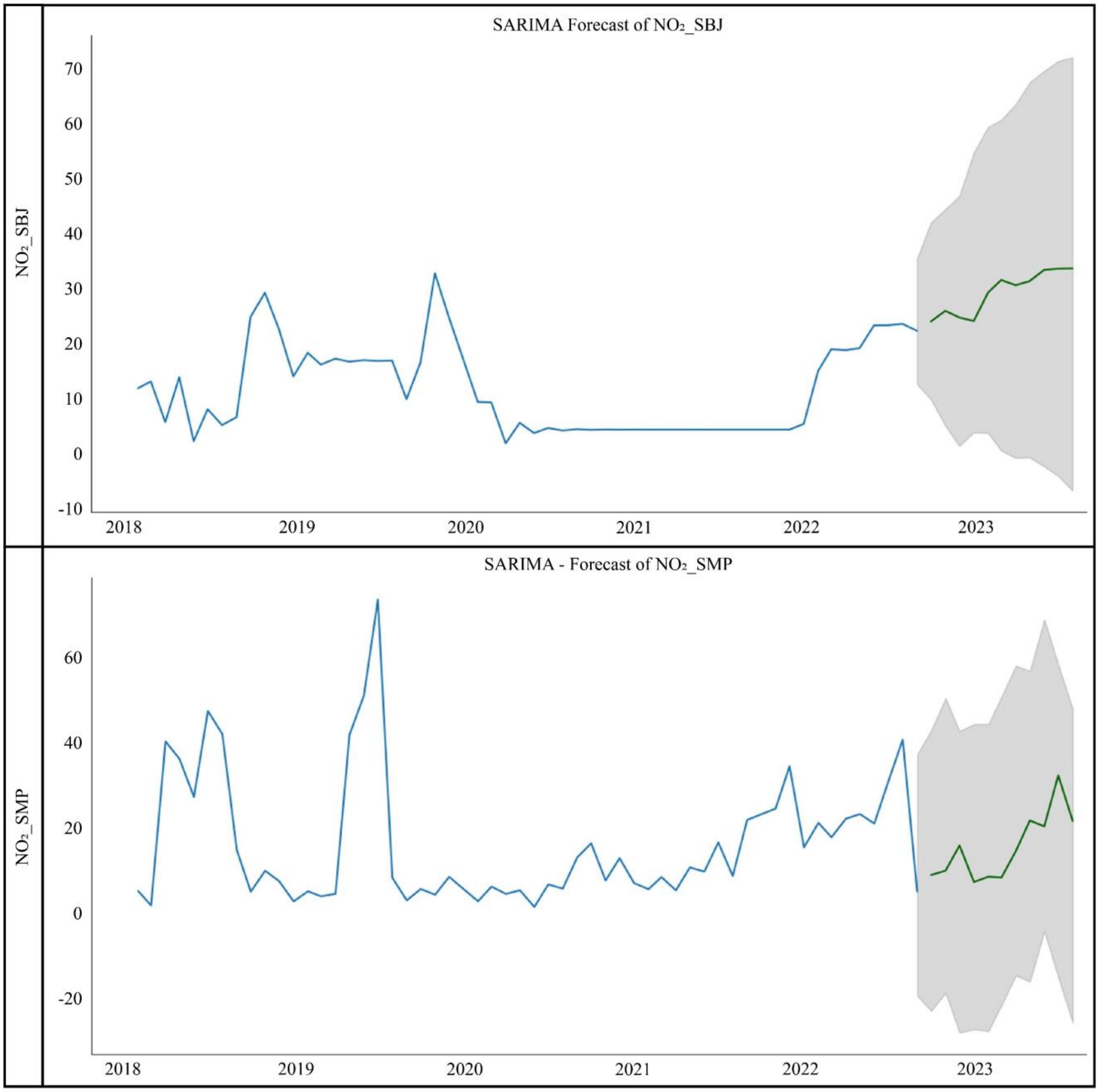


**Fig. 5 $NO_2$ Forecast for both districts**

**Table 6. $NO_2$ forecast results for both districts**

| Month | Year | $NO_2$ Forecast | |
|---|---|---|---|
| | | **San Borja** | **San Martin de Porres** |
| September | 2022 | 23.7 | 8.55 |
| October | 2022 | 25.65 | 9.57 |
| November | 2022 | 24.44 | 15.44 |
| December | 2022 | 23.79 | 6.92 |
| January | 2023 | 29.01 | 8.15 |
| February | 2023 | 31.29 | 7.95 |
| March | 2023 | 30.34 | 14.16 |
| April | 2023 | 31.08 | 21.3 |
| May | 2023 | 33.13 | 19.91 |

### 4.3. SARIMA Model Forecast

Forecasts were made with the SARIMA model in both districts; Figure 4 and Figure 5 shows the results. Comparing the concentrations of both communities, before the pandemic and during the years 2018 and 2019, the district of San Martin de Porres presented higher concentrations; this may be because the district of San Borja offers a high loss of data in several months of these years and the imputation method used was not the most accurate for the number of months. Likewise, this problem is repeated for the years 2020 and 2021; this problem is repeated since the amount of missing data in the district of San Borja is high compared to the data collected from the San Martin de Porres district.

### 4.4. Limitations and Future Scope of Work

This research faces several significant challenges. The lack of data on $NO_2$ concentration measured by SENAMHI has required data imputation, which must be clearly detailed. In addition, the focus on only two districts of Lima limits the ability to generalize the results to the entire metropolis. The lack of previous research employing time series models to forecast $NO_2$ concentration in Lima districts makes comparison with relevant literature difficult. In addition, it is important to recognize the inherent limitations of time series models, such as the assumption of nonlinear seasonality and sensitivity to outliers, which could affect the accuracy of the results. Despite these challenges, this research seeks to contribute to the existing knowledge on air quality in Lima and its implications. It is also recommended that for possible future research, it is recommended to collect large amounts of data covering a longer period than the one currently evaluated in this article to consider those exogenous variables that could improve the calculation of the $NO_2$ concentration forecast and to extend the research to more districts of Metropolitan Lima.

## 5. Conclusion

Based on the information available from SENAMHI between 2018 and 2022, the average $NO_2$ concentration between both districts was relatively low, with the value of the San Martin de Porres district being slightly higher. The Dickey-Fuller test compared the time series and verified seasonality and non-seasonality. The AIC and BIC criteria helped them find the best model for their time series. On the other hand, the ARIMA model showed that it could not make an accurate forecast; therefore, it is necessary to test different options, such as the SARIMA model, which, according to this investigation, could accurately forecast the $NO_2$ concentration for the following nine months. of the districts of San Borja and San Martin de Porres. The importance of controlling and monitoring the concentration of $NO_2$ in the air can be highlighted. This implies that the environmental authorities must develop policies and strategies that help maintain $NO_2$, mainly in the San Borja district, due to the evidence of higher concentrations of $NO_2$ in the future. These findings are relevant to the Sustainable Development Goals (SDGs) since it is closely related to SDG 3 (Health and Well-being) since air pollution by $NO_2$ negatively impacts human health, such as respiratory and cardiovascular diseases. Likewise, it is related to SDG 11 (Sustainable Cities and Communities) because poor air quality does not guarantee the proper development of sustainable cities and communities.

Another important point to highlight about the concentration of $NO_2$ in San Borja and San Martin de Porres is the natural effect of dispersion, transformation, and removal of pollutants in each of these districts; this is very complex since it depends on factors such as precipitation, relative humidity, air temperature, radiation, wind speed, and direction. The wind direction in Metropolitan Lima goes with more intensity and direction from north to south; San Martin de Porres is in North Lima, and San Borja is in South Lima. In addition, the automobile sector, one of the most polluting in the country, is in Lima Center, where emissions exceed Environmental Quality Standards (ECA). Consequently, the air masses mentioned in the previous point play a crucial role in the contamination of the area. It should be noted that the forecast results and current values of the $NO_2$ concentration are in the healthy range of ECA.

## References


[1] Faqih Hamami, and Inayatul Fithriyah, “Classification of Air Pollution Levels using Artificial Neural Network,” *International Conference on Information Technology Systems and Innovation*, pp. 217-220, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] Nishant Raj Kapoor et al., “Machine Learning-Based $CO_2$ Prediction for Office Room: A Pilot Study,” *Wireless Communications and Mobile Computing*, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Gonzalo Carreño et al., “Machine Learning Models to Predict Critical Episodes of Environmental Pollution for PM2.5 and PM10 in Talca, Chile,” *Mathematics*, vol. 10, no. 3, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] Ahmed Alnaim, Ziheng Sun, and Daniel Tong, “Evaluating Machine Learning and Remote Sensing in Monitoring $NO_2$ Emission of Power Plants,” *Remote Sensing*, vol. 14, no. 3, pp. 729, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Pratima Kumari, and Durga Toshniwal, “Impact of Lockdown on Air Quality Over Major Cities Across the Globe During the COVID-19 Pandemic," *Urban Climate*, vol. 34, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] Yanding Wang et al., “Prediction and Analysis of COVID-19 Daily New Cases and Cumulative Cases: Times Series Forecasting and Machine Learning Models,” *BMC Infectious Diseases*, vol. 22, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Erick Chávez Flores, “Incidence of Covid-19 Quarantine on Air Quality ($NO_2$) in the City of Lima,” *Journal of Research Institute of the Faculty of Mines, Metallurgy and Geographical Sciences*, vol. 23, no. 46, pp. 65-71, 2020. [CrossRef] [Publisher Link]

[8] Vilma Tapia et al., "Vehicle Rearrangement and Environmental Pollution by Particulate Matter (2.5 And 10), Sulfur Dioxide and Nitrogen Dioxide in Lima Metropolitana, Peru," *Peruvian Journal of Experimental Medicine and Public Health*, vol. 35, no. 2, pp. 190-197, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[9] Jaime Yelsin Rosales Malpartida, "Prediction of Air Pollution Generated by $CO_2$ Emissions in Peru using ARIMA Methods and Neural Networks," *Journal BioFab*, vol. 2, no. 1, pp. 130-142, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Raghavendra Kumar, Pardeep Kumar, and Yugal Kumar, "Time Series Data Prediction using IoT and Machine Learning Technique," *Procedia Computer Science*, vol. 167, pp. 373-381, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[11] Pooja Gopu et al., "Time Series Analysis Using ARIMA Model for Air Pollution Prediction in Hyderabad City of India," *Soft Computing and Signal Processing*, vol. 1325, pp. 47-56, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Stylianos I. Vagropoulos et al., "Comparison of SARIMAX, SARIMA, Modified SARIMA, and ANN-Based Models for Short-Term PV Generation Forecasting," *IEEE International Energy Conference*, pp. 1-6, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[13] Norizan Mohamed et al., "Improving Short Term Load Forecasting Using Double Seasonal Arima Model," *World Applied Sciences Journal*, vol. 15, no. 2, pp. 223-231, 2011. [Google Scholar] [Publisher Link]

[14] K.Krishna Samal et al., "Time Series Based Air Pollution Forecasting using SARIMA and Prophet Model," *ITCC 2019: Proceedings of the 2019 International Conference on Information Technology and Computer Communications*, pp. 80-85, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[15] Francesca Lazzeri, *Machine Learning for Time Series Forecasting with Python*, 1st Edition., Wiley, pp. 22, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[16] Esperanza Manrique Rojas, "Machine Learning: Analysis of Programming Languages and Development Tools," *Iberian Journal of Information Systems and Technologies*, vol. 4, no. 28, pp. 586-599, 2020. [Publisher Link]

[17] Małgorzata Murat et al., "Forecasting Daily Meteorological Time Series using ARIMA and Regression Models," *International Agrophysics*, vol. 32, no. 2, pp. 253-264, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[18] Gonzalo Ríos, and Carlos Hurtado, "Time Series," *Topics in Data Mining*, 2008. [Publisher Link]

[19] Liming Ye et al., "Time-Series Modeling and Prediction of Global Monthly Absolute Temperature for Environmental Decision Making," *Advances in Atmospheric Sciences*, vol. 30, no. 2, pp. 382-396, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[20] George E.P. Box et al., *Time Series Analysis: Forecasting and Control*, 5th Edition, Ed. Hoboken, New Jersey: Wiley, vol. 37, no. 5, pp. 712, 2015. [Google Scholar] [Publisher Link]

[21] Shreya Kusrey, Avinash Rai, and Vineeta (Nigam) Saxena, "Zigbee Based Air Pollution Monitoring and Control System using WSN," *SSRG International Journal of Electronics and Communication Engineering*, vol. 4, no. 6, pp. 7-11, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[22] Patrima Kumari, and Durga Toshniwal, "Impact of Lockdown on Air Quality Over Major Cities across the Globe During the COVID-19 Pandemic," *Urban Climate*, vol. 34, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[23] Fahad Radhi Alharbi, and Dénes Csala, "A Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) Forecasting Model-Based Time Series Approach," *Inventions*, vol. 7, no. 4, pp. 94, 2022. [CrossRef] [Google Scholar] [Publisher Link]