AIMS Environmental Science

http://www.aimspress.com/journal/environmental

AIMS Environmental Science, 11(3): 401–425
DOI: 10.3934/environsci.2024020
Received: 28 November 2023
Revised: 02 April 2024
Accepted: 09 May 2024
Published: 31 May 2024

Research article

Multi-step ahead ozone level forecasting using a component-based technique: A case study in Lima, Peru

Flor Quispe[1], Eddy Salcedo[1], Hasnain Iftikhar[2], Aimel Zafar[3,4], Murad Khan[5], Josué E. Turpo-Chaparro[6], Paulo Canas Rodrigues[7], Javier Linkolk López-Gonzales[6,*]

[1] E.P. Ingeniería Ambiental, Universidad Peruana Unión, Lima, Peru
[2] Department of Statistics, Quaid-i-Azam University, 45320, Islamabad, Pakistan
[3] Department of Statistics, University of Peshawar, Pakistan
[4] Department of Mathematics, Statistics and Computer Science, The University of Agriculture, Peshawar - Pakistan
[5] Department of Statistics, Abdul Wali Khan University Mardan, Mardan 23200, Pakistan
[6] Escuela de Posgrado, Universidad Peruana Unión, Lima 15468, Peru
[7] Department of Statistics, Federal University of Bahia, Salvador 40170-110, Brazil

* **Correspondence:** javierlinkolk@gmail.com

Abstract: The rise in global ozone levels over the last few decades has harmed human health. This problem exists in several cities throughout South America due to dangerous levels of particulate matter in the air, particularly during the winter season, making it a public health issue. Lima, Peru, is one of the ten cities in South America with the worst levels of air pollution. Thus, efficient and precise modeling and forecasting are critical for ozone concentrations in Lima. The focus is on developing precise forecasting models to anticipate ozone concentrations, providing timely information for adequate public health protection and environmental management. This work used hourly O_3 data in metropolitan areas for multi-step-ahead (one-, two-, three-, and seven-day-ahead) O_3 forecasts. A multiple linear regression model was used to represent the deterministic portion, and four-time series models, autoregressive, nonparametric autoregressive, autoregressive moving average, and nonlinear neural network autoregressive, were used to describe the stochastic component. The various horizon out-of-sample forecast results for the considered data suggest that the proposed component-based forecasting technique gives a highly consistent, accurate, and efficient gain. This may be expanded to other districts of Lima, different regions of Peru, and even the global level to assess the efficacy of the proposed component-based modeling and forecasting approach. Finally, no analysis has been undertaken using a component-based estimation to forecast ozone concentrations in Lima in a multi-step-ahead manner.

Keywords: multi-step ahead ozone forecasting; global health; multiple linear regression model; time series models; component-based forecasting technique

1. Introduction

Air pollution is a major issue in the modern industrialized world which has severe toxicological effects on human health and the environment [1]. The study conducted by Monash University in Australia found that the World Health Organization's recommended criteria for air quality were not being met where 99% of the world's population resided [2]. Even though various physical activities release pollutants, unintentionally releasing hazardous chemicals is the primary cause of pollution. Elevated tropospheric ozone (O_3 ($\mu g/m^3$)) concentrations indicate a major hazard to the climate and environment. Additionally, the dispersion of O_3 is hampered by climate change due to industrial activities and urbanization [3]. Important air pollution indicators include Nitrogen Dioxide (NO_2), O_3, the absorbing aerosol index (AAI), and carbon monoxide (CO). The main contributors to atmospheric NO_2 production include soil emissions, natural lightning, motor vehicle exhausts, biomass burning, and partial combustion of fossil fuels. It is crucial in synthesizing tropospheric ozone through intricate chemical processes involving oxygen and free radicals produced by sunlight on volatile organic compounds (VOCs) [4]. Since O_3 is produced as a by product of the photochemical reaction between CO and VOC and nitrogen oxides (NOx = NO + NO_2), which enables its high concentrations to be produced by NOx emissions from combustion sources, O_3 is regarded as a secondary pollutant [6]. The World Health Organization states that epidemiological and toxicological investigations have found significant support for the causal relationship between surface O_3 and unfavorable respiratory consequences. Especially in populations at high risk, these impacts can include mortality as well as changes to lung function and asthma [5]. Ozone harms crops as well as plant foliage [8]. To inform the public about the necessary intervention and to assess the immediate effects of activities on climate behavior, it is vital to anticipate and comprehend the rate of ozone generation and emission [7].

Globally, long-term ozone exposure was projected to contribute to an additional 254,000 chronic obstructive pulmonary disease deaths [12, 13]. China is one of the nations with the highest ozone emissions and concentrations on a worldwide scale [7, 9]. Beijing and Shanghai have had the worst air pollution in recent years, with the critical days of O_3 pollution being 93 to 575% greater than those of other industrialized nations [9]. In contrast, some regions in the United States and southern Canada experience less ozone exposure and are called "clean places" [14]. The metropolitan area of Mexico City, the country's capital, has 21.8 million residents and is in a high-altitude basin (about 2240 m above sea level). Mexico City is frequently subject to ozone episodes because of its special topographical environment and meteorological and emission conditions. These events seem to have lately gotten worse again [10]. However, because tropical and subtropical areas have favorable climatic conditions for ozone production and accumulation, such as high temperatures, intense sunlight, and convection, the variation in emissions from these regions can be seen in the global ozone load, demonstrating the close connection between climatic factors and O_3 concentration [11]. Improvements in ozone air quality, particularly in Europe and North America, have been made by reducing anthropogenic emissions of ozone precursors like nitrogen oxides (NOx). Peru is included among the nations with the highest levels of air pollution. This condition is linked to Peru's quick economic and industrial development,

which results in the production of pollutants and gases that affect the quality of the air. Lima has over a third of the nation's population as the capital of Peru, making it the city with the greatest air pollution in South America [16].

Statistical modeling approaches have been widely applied to air pollution to describe the interactions between variables. A relationship between many explanatory factors (predictors) and a response variable (target) is typically established using statistical techniques based on regression models, such as multiple linear regression (MLR). But, despite their apparent ability to deliver reasonable results in many applications, these types of models frequently fall short in describing the complexity of non-linear relationships and interactions between variables, and more advanced techniques are typically preferred to achieve a higher degree of accuracy in predictions of pollutant concentration levels [19, 20]. The literature has provided several statistical methods for forecasting and evaluating ozone pollution levels, including the autoregressive model, the autoregressive integrated moving average, and its model variants [21, 22, 23]. Scientific research is currently using machine learning (ML) models more and more. Due to their ability to analyze vast and complex datasets (big data), find patterns, and make predictions, advanced statistical models based on ML approaches have been created and used in the field of air quality modeling more and more during the past three decades [17, 18]. During the COVID-19 outbreak in Spain, the issue with NO_2 was addressed using ML techniques [24]. Using several methods, a study was done by [25] to forecast Jordan's ground-level ozone concentrations. They discovered that an algorithm based on artificial neural networks performed better than all other methods. The study was done to forecast hourly ozone concentrations for the next day using a novel approach based on feedforward artificial neural networks with principal components as inputs. The multiple linear regression and feedforward artificial neural networks were compared to the developed model based on the original data and using principal component regression. The results revealed that using principal components as inputs improved both models' predictions by reducing their complexity and eliminating data collinearity [26]. In an empirical investigation, [27] used a standard support vector machine (SVM) to forecast ozone levels based only on environmental factors. The outcomes showed that the SVM performed better than neural networks in forecasting daily maximum ozone concentrations. To model ozone concentrations throughout the continental United States, [28] evaluated thirteen ML techniques with linear land-use regression (LUR). The nonlinear ML techniques outperformed LUR regarding prediction accuracy, with the improvement being more significant for spatiotemporal modeling. By adjusting the sample weights, spatiotemporal models can anticipate concentrations needed to determine ozone design values that are as good as or better than spatial models. The aim of the study by [29] was to predict tropospheric (O_3) using a dataset of ozone concentrations using a variety of ML models, including linear regression, tree regression, support vector regression, ensemble regression, Gaussian process regression, and artificial neural network models. For the prediction of ozone pollution, [30] assessed the predictive effectiveness of 19 ML algorithms. According to the findings, dynamic ML models that use time-lagged data perform better than static and reduced ML models. When comparing ML models to static and reduced models, time-lagged data increases accuracy by 300% and 200%, respectively, according to RMSE measures.

Peru is a South American country in the Southeast Pacific Region, and its capital, Lima, is no stranger to ozone air pollution. Lima has grown into a megacity with over ten million people and severe air pollution concerns. Romero et al. [45] investigated the impact of meteorological variables on ozone concentrations and other pollutants in the air using linear correlations for data collected between

2015 and 2018 at eight different sampling stations in metropolitan Lima and found that this pollutant increased with solar irradiation between 10:00 and 16:00 hours, particularly in spring, possibly due to the interaction of primary NOx and hydrocarbon emissions from vehicles. Carbo-Bustinza et al. [11] instead investigated the behavior of ozone in winter at four sites in Lima using ML techniques and discovered the most significant critical values in the Ate region. However, they detected a general decrease in values during the cold season (100 $\mu g/m^3$), consistent with another study [46]. Meanwhile, there is a requirement to thoroughly analyze the time series of the most contaminated areas to optimize the O_3 forecast.

In this respect, this study aims to provide an improved tool for forecasting tropospheric ozone concentrations in four districts of the megacity of Lima using a components-based estimate approach. In a highly accurate and efficient manner, a component-based technique combines the features of classical multiple regression models and time series models to create efficient forecasts. This study made the following contributions: To improve the efficiency and accuracy of O_3 forecasting, a component-based forecasting technique based on the multiple linear regression model and four standard time series models is proposed. The application of the component-based forecasting technique of the O_3 database in four districts: Ate, Campo de Marte (CDM), San Borja (SB), and Santa Anita (STA), with severe episodes of ozone contamination between 2017 and 2019 only for the winter season. Six different accuracy mean errors were used to evaluate the performance of the proposed component-based forecasting technique, including three relative and three absolute accuracy mean errors, a statistical test, and a visual evaluation. In addition, four different forecast horizons are used to evaluate the short- to medium-term forecasting performance. On the other hand, in this work, the results of the final best model are compared with the considered baseline models. The findings showed that the best model in this study is highly accurate and efficient compared to the benchmark models. Likewise, a methodological proposal applicable to the environmental management system to mitigate ozone pollution is provided, aimed at the stakeholders of the national air quality program. Finally, the current work uses only four district datasets in Lima, Peru. This can be extended to other districts of Lima, other regions of Peru, and even the world level to evaluate the performance of the proposed component-based forecasting technique. Finally, no analysis has been undertaken using a component-based estimation to forecast ozone concentrations in Lima in a multi-step-ahead manner.

This research was motivated by the urgent worldwide air pollution problem and its significant environmental and human health effects. Peru receives special attention because of its fast industrial expansion and high levels of air pollution, especially in Lima. Recognizing the limits of standard statistical models, the work offers an enhanced component-based forecasting approach to increase accuracy in predicting tropospheric ozone concentrations. The main objective is to offer a useful instrument for managing the environment, intervening when necessary, and maybe being used globally to lessen the negative consequences of air pollution. The remaining manuscript is formatted as follows: The proposed component-based forecasting approach is explained in detail in Section 2. Section 3 contains the outcomes of the case studies for each monitoring station analyzed and some meaningful discussion. Section 4 presents the results, limits, and future challenges.

2. Method and materials

This section presents the study area, the distribution of the monitoring stations, and the data sources used. Likewise, this section will comprehensively overview the various models and methods used to construct the proposed component-based modeling and forecasting technique. Thus, the subsequent subsections provide detailed information on each model and method.

2.1. Data understanding

This work uses hourly O_3 datasets from four monitoring stations in the Lima metropolitan area (see Figure 1): Ate, CDM, SB, and STA, for three consecutive years, 2017, 2018, and 2019. Only the winter days of each year are considered. As a result, for one station, there are 6768 data points: a training section (for model fit) and a testing section (for out-of-sample forecast). The training section comprises data from 2017 to 2018, the first two years (4512 hours), while the one the complete year of 2019 (2256 hours) is utilized as out-of-sample data (testing). It is common practice to prepare the data before beginning the modeling process. The purpose of preprocessing is generally to make data modeling easier. To do this, the database is sorted, categorized, and evaluated for each monitoring station while accounting for the city's winter season, which spans from June 21 to September 22, for ozone. From 2017 to 2019, four monitoring stations were proposed at essential places in Lima, Peru's capital. It should be mentioned that the capital, Lima, has ten monitoring stations; nevertheless, four were chosen owing to a lack of data in the registration. A Teledyne analyzer was used to test the ozone concentrations every hour. Zero and span testing, calibration, and leak detection are all examples of analyzer activities. After correcting zeros, duplicates, and/or anomalies, the data is relayed via telemetry to *Servicio Nacional de Meteorología e Hidrología del Perú* (SENAMHI) for certification. Similarly, SENAMHI features a systematic network of stations that monitor and report the variables investigated to a processing center regularly and automatically. On an hourly basis, these stations employ high-quality instrumentation and sensors to detect temperature, relative humidity, wind speed, and direction. Furthermore, an inductive approach, Multiple Imputation by Chained Equations, was used. This approach is built on an utterly conditional specification, with each incomplete variable given by its model [38].

2.2. The proposed component-based modeling and forecasting technique

The primary goal of this study was to predict the O_3 level one, two, three, and seven days ahead at four monitoring stations: Ate, CDM, SB, and STA in Lima, Peru. Let O_h represent the O_3 for the h^{th} hour. To accurately account for the changes in O_3 over time, we suggest modeling O_h in the following way:

$$O_h = d_h + s_h \tag{2.1}$$

The ozone concentration series is split into two parts: a deterministic component (denoted as d_h) and a stochastic component (denoted as s_h). d_h includes the trend (long-term pattern) and hourly cycles, while s_h represents random fluctuations. Mathematically, d_h is defined as follows:

$$d_h = t_h + n_h \tag{2.2}$$

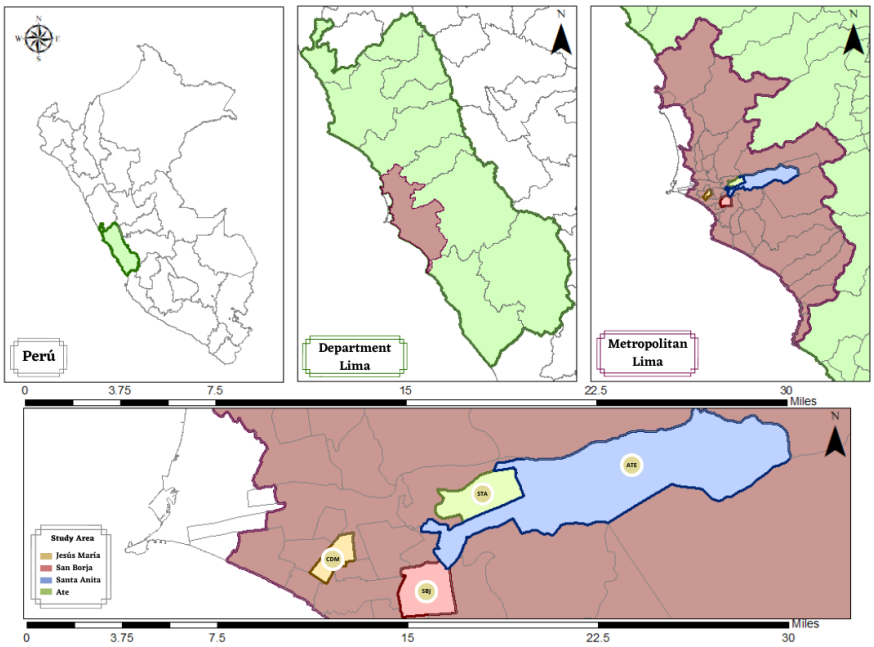

Figure 1. Map with the metropolitan area of Lima, Peru, together with the location of the four pollutant and weather monitoring stations that belong to SENAMHI: Ate, CDM, SB, and STA.

The symbol t_h represents the long-term trend, while n_h represents the hourly periodicity component. On the other hand, s_h is a stochastic component, also known as residuals, that defines the random dynamics. A multiple linear regression model estimates the deterministic component d_h. To estimate stochastic components, this study examines four distinct models for univariate time series analysis: autoregressive, nonparametric autoregressive, autoregressive moving average, and nonlinear neural network autoregressive. As a result, there are four possible combinations for comparison purposes when deterministic and stochastic models are combined.

2.2.1. Procedure for modeling the deterministic component

This section will discuss estimating the deterministic component using a multiple linear regression model. To achieve this, we will model the response variable O_h linearly by estimating the trend (long-run) component t_h through linear regression for time h. Additionally, we will describe the hourly periodicity using dummies:$n_h = \sum_{i=1}^{24} \zeta_i I_{i,h}$. The variable $I_{i,h}$ is assigned a value of 1 when h refers to the i^{th} hour of the day and 0 otherwise. The regression coefficients (ζ_i) associated with these components are determined using the ordinary least square method. After obtaining all the regression coefficients, the estimated trend and hourly periodicity equation are presented.

$$\hat{d}_h = \hat{\zeta}_0 t_h + \sum_{i=1}^{24} \hat{\zeta}_i I_{i,h}; \tag{2.3}$$

Once the estimated deterministic component is obtained, the residual or stochastic component can

be derived as

$$s_h = O_h - (\hat{d}_h) \tag{2.4}$$

2.3. Modeling the stochastic component

The residual series was obtained from both models using a multiple linear regression model to estimate the stochastic component in this work. However, to model and forecast the stochastic component, four different univariate time series models are considered: autoregressive, nonparametric autoregressive, autoregressive moving average, and nonlinear neural network autoregressive models [31, 32]. Details on these models are given in the following section.

2.3.1. Autoregressive model

A linear and parametric autoregressive (AR) process describes the short-term dynamics of s_h and considers a linear combination of the previous time d observations of s_h, denoted as

$$s_h = \alpha + \vartheta_1 s_{h-1} + \vartheta_2 s_{h-2} + ... + \vartheta_n s_{h-n} + \varepsilon_h \tag{2.5}$$

In the above formula, α is an intercept term, $\vartheta_j (j = 1, 2, \cdots, n))$ is the slope parameter of the underlying AR process, and ε_t is the disturbance term. The most appropriate form of the AR (n) model is the following: This model is defined by one parameter, say (n). The s represents the number of past observations used in the model and captures the influence of past data points on the current value. However, the AR model order selection is established by inspecting the correlograms (i.e., ACF and PACF). This work fits the AR (5), AR (3), AR (4), and AR (5) models for Ate, CDM, SB and STA, respectively.

2.3.2. Autoregressive moving average model

The autoregressive moving average (ARMA) model incorporates the target variable's past values and utilizes important information as moving average(s). In our case, the study variable s_h is explained on the previous n terms, as well as the lagged values of residuals. Mathematically,

$$s_h = \alpha + \vartheta_1 s_{h-1} + \vartheta_2 s_{h-2} + \cdots + \vartheta_n s_{h-n} + \varepsilon_h + \zeta_1 \varepsilon_{h-1} + \zeta_2 \varepsilon_{h-2} + ... + \zeta_m \varepsilon_{h-m} \tag{2.6}$$

In the last equation, α denotes the intercept, $\vartheta_j (j = 1, 2, \cdots, n)$ and $\zeta_k (k = 1, 2, \cdot, m)$ are the parameters of AR and MA process respectively, and ε_h is a Gaussian white noise series with mean zero and variance σ_ε^2. The ARMA (n,m) model is defined by two parameters: n and m. The parameter n represents the number of past observations used (AR order), while the parameter m represents the number of past forecast errors included (MA order). The AR component shows the influence of past data points on the current value, while the MA component accounts for the impact of past forecasting errors. This study inspects the correlograms (i.e., ACF and PACF) to select the ARMA model order. In this work, we fit the ARMA (5,2), AR (3,2), AR (4,3), and AR (5,1) models for the Ate, CDM, SB, and STA, respectively.

2.3.3. Nonparametric Autoregressive Model

The additive nonparametric counterpart of the AR process leads to the additive model (NPAR), where the association between s_h and its previous terms do not have any specific parametric form,

which is stated as

$$s_h = q_1(s_{h-1}) + q_2(s_{h-2}) + \ldots + q_n(s_{h-n}) + \varepsilon_h \tag{2.7}$$

where $q_j(j = 1, 2, \cdots, n)$ are smoothing functions and describe the association between s_h and its previous values. In this work, the functions q_i are denoted by cubic regression splines. As in the case of the parametric AR form, considered the NPAR (5), NPAR (3), NPAR (4), and NPAR (5) models for the Ate, CDM, SB, and STA, respectively.

2.3.4. Autoregressive veural network

An autoregressive neural network (NNA) is a machine learning model that predicts the values of input variables in the future. The NNA model predicts future values of a time series s_h based on its past observations, such as the mathematical function given by $s_{h-1}, s_{h-2}, ..., s_{h-n}$ [33]. In this expression, n is the time delay parameter. The NNA model is trained using the backpropagation method and the steepest descent approach to reduce the squared error between the actual and predicted values. The NNA (n,m) model is a suitable artificial neural network form that relies on n and m parameters. In this model, n represents the number of past observations (nodes), while m represents the number of hidden layers (delayed input). In this study, we utilized the NNA (5,3), NNA (3,2), NNA (4,2), and NNA (5,2) models for Ate, CDM, SB, and STA, respectively.

In addition to the above-stated models, we include two baseline models, the Naive and the Seasonal Naive models, to assess the performance of the proposed component-based forecasting models. The details about the baseline models are given by

2.3.5. The naive model

One of the most basic time series forecasting models is the naïve forecast, frequently used as a benchmark for evaluating the effectiveness of other techniques. It merely makes the best estimate of the future value based on the variable's most recent value [47]. That is,

$$\hat{O}_{T+h—T} = O_T \tag{2.8}$$

2.3.6. Seasonal naive model

For the seasonal data, a similar approach, the seasonal naïve method, helps forecast time series analysis. In this situation, each forecast is made to equal the most recent observation from the same season (for example, the same hour or day of the week or the month of the year within the seasonal dataset). In a formal setting, the forecast for time T + h is expressed as

$$\hat{O}_{T+h—T} = O_{T+h-m(k+1)} \tag{2.9}$$

where k is the integer portion of $(h-1)/m$ (i.e., the number of full years in the forecast period previous to time T+h), and m is the seasonal period.

Thus, once both deterministic and stochastic components are forecasted using the respective models, the final one to seven days ahead forecasts are derived as

$$\hat{O}_{h+1} = (\hat{t}_{h+1} + \hat{n}_{h+1} + \hat{s}_{h+1}) \tag{2.10}$$

It is worth mentioning here that the proposed component-based modeling and forecasting technique is motivated by the following literature work [34, 35, 36, 37].

2.4. Accuracy measures

Six different standard accuracy measures were calculated to validate the performance of the proposed component-based modeling and forecasting technique, including mean absolute error (MAE), mean absolute percentage error (MAPE), symmetric mean absolute percentage error (SMAPE), root mean square error (RMSE), root means squared log error (RMSLE), and root relative squared error (RRSE) [39]. The MAE, MAPE, SMAPE, RMSE, RMSLE, and RRSE formulae are shown below:

$$\text{MAE} = \frac{1}{H}\sum_{h=1}^{H}|O_h - \hat{O}_h|, \tag{2.11}$$

$$\text{MAPE} = \frac{1}{H}\sum_{h=1}^{H}\left|\frac{O_h - \hat{O}_h}{O_h}\right|, \tag{2.12}$$

$$\text{SMAPE} = \frac{1}{H}\sum_{h=1}^{H}\frac{|O_h - \hat{O}_h|}{(|O_h| + |\hat{O}_h|)/2}, \tag{2.13}$$

$$\text{RMSE} = \sqrt{\sum_{h=1}^{H}\frac{(O_h - \hat{O}_h)^2}{H}}, \tag{2.14}$$

$$\text{RMSLE} = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(log(O_h + 1) - log(\hat{O}_h + 1))^2}, \tag{2.15}$$

$$\text{RRSE} = \sqrt{\frac{\sum_{h=1}^{H}(O_h - \hat{O}_h)^2}{\sum_{h=1}^{H}(O_h - \bar{O})^2}} \tag{2.16}$$

In the above equations, O_h is observed and $\hat{O}_h$ is the forecasted ozone value for h^{th} observation (h=1, 2, . . ., 2256=H).

In addition to accuracy performance measures, to assess the significance of the differences in the prediction performance of the proposed models, the Diebold-Mariano test was performed [40]. The DM test is a widely used statistical test for comparing predictions obtained from different models [41, 42, 43]. The DM statistic is given by

$$DM_s = \frac{\bar{X}}{\sqrt{Var(\bar{X})}} \tag{2.17}$$

where

$$\bar{x} = \frac{1}{H}\sum_{h=1}^{H}X_h,\ X_h = (O_h - \tilde{O}_{1h})^2 - (O_h - \tilde{O}_{2h})^2, \tag{2.18}$$

$$\mathrm{Var}(\bar{\mathrm{X}}) = \frac{1}{\mathrm{H}}(2\sum_{j=1}^{\mathrm{h}-1} r_j + r_0), \ \textit{and} \ r_j = \mathrm{cov}(\mathrm{X_h} - \mathrm{X_h} - j). \tag{2.19}$$

$\tilde{\mathrm{O}}_{1\mathrm{h}}$ is the predicted value of the first predictive model, and $\tilde{O}_{2\mathrm{h}}$ is the predicted value of the second predictive model at time h. If the DM statistic is negative, the first predictive model is statistically better than the second predictive model.

3. Case study evaluation and discussion

To obtain the forecasts for the O_3 concentration one day ahead, two days ahead, three days ahead, and seven days ahead, using the proposed component-based methodology for time series forecasting presented in Section 2 for all considered monitoring stations, the following steps need to be followed: First, to stabilize the variance of the O_3 concentration time series, the natural logarithmic transformation was applied. Second, we divided the hourly O_3 concentration time series into two new components: deterministic and stochastic. The deterministic component contains a linear long-trend component and an hourly seasonal component, while the stochastic is the remainder. To model the deterministic part using a multiple linear regression model and the stochastic component with various time series models discussed in the last section. Finally, both components' forecasts are combined to get the final forecast results for each possible combination model. Therefore, the forecasts of one day ahead, two days ahead, three days ahead, and seven days ahead were obtained using the expanding window technique for 94 days (2256 hours), and the models were estimated accordingly. Likewise, the O_3 forecasts were achieved through equation 2.10. The performance measures, including MAE, MAPE, SMAPE, RRAE, RMSLE, and RMSE, are then used for the evaluation and comparative performance of the models. Therefore, the details of the results from four monitoring stations are given in the following tables: Ate (in Table 1), CDM (in Table 2), SB (in Table 3) and STA (in Table 4), all located in metropolitan Lima, Peru.

The results of the Ate are listed in Table 1. This table shows the accuracy mean errors of the four horizons, such as one day, two days, three days, and seven days ahead, of the following six models: four combination models from within the proposed component-based forecasting technique: the AR, the NPAR, the ARMA, and the NNA models; and two baseline models: the naive and seasonal naive models. The following two conclusions were drawn from Table 1. The mean accuracy errors of the NPAR model were minimal. As shown in Table 1, the NPAR model had the best forecasting effect, with accuracy mean errors (MAEs, MAPEs, SMAPEs, RRSE, RMSLE, and RMSE) of the one-day (2.692, 0.192, 0.194, 1.022, 0.229, and 3.685), two-day (2.879, 0.216, 0.217, 1.116, 0.309, and 3.990), and seven-day (2.900, 0.205, 0.223, 1.117, 0.309, and 3.999) ahead forecasts, respectively, less than the AR, ARMA, and NNA models within the proposed forecasting methodology, and also significantly minimal to the baseline models (the naive and seasonal naive models). The predictive effect of the NNA model was the worst and was much higher than the mean errors of the AR and ARMA models. However, only in the case of three-day forecast accuracy mean errors are the best results shown by the ARMA models with the following metrics: MAE = 2.879, MAPE = 0.216, SMAPE = 0.217, RRAE = 1.116, RMSLE = 0.309, and RMSE = 3.990. Although the NPAR model also shows the second-best results, on the other hand, comparing the best model within the proposed forecasting approach with the baseline models (naive and season-naive models), it is confirmed from Table 1 that the NPAR model outperforms the

Table 1. The O_3 in Ate Station: The mean forecast error for all models for a twenty-four-hour ahead out-of-sample forecast.

One-day-ahead (24 hours ahead)						
MODEL	MAE	SMAPE	MAPE	RRSE	RMSLE	RMSE
AR	2.997	0.212	0.218	1.113	0.247	4.010
NPAR	2.692	0.192	0.194	1.022	0.229	3.685
ARMA	2.790	0.199	0.200	1.043	0.233	3.760
NNA	3.276	0.224	0.228	1.464	0.286	5.276
NAÏVE	3.181	0.219	0.231	1.219	0.263	4.394
SNAÏVE	3.829	0.277	0.289	1.391	0.326	5.012
Two-day-ahead (48 hours ahead)						
MODEL	MAE	MAPE	SMAPE	RRSE	RMSLE	RMSE
AR	3.132	0.231	0.224	1.174	0.262	4.192
NPAR	2.699	0.194	0.190	1.073	0.237	3.832
ARMA	2.823	0.202	0.203	1.078	0.239	3.849
NNA	3.452	0.245	0.230	1.578	0.308	5.635
NAÏVE	3.467	0.255	0.237	1.360	0.288	4.855
SNAÏVE	3.792	0.287	0.275	1.392	0.324	4.971
Three-day-ahead (72 hours ahead)						
MODEL	MAE	MAPE	SMAPE	RRSE	RMSLE	RMSE
AR	3.112	0.241	0.231	1.219	0.326	4.361
NPAR	2.944	0.218	0.217	1.185	0.316	4.237
ARMA	2.879	0.216	0.218	1.116	0.309	3.990
NNA	3.940	0.294	0.263	1.812	0.391	6.479
NAÏVE	3.769	0.297	0.265	1.523	0.371	5.446
SNAÏVE	3.853	0.292	0.288	1.415	0.372	5.060
Seven-day-ahead (168 hours ahead)						
MODEL	MAE	MAPE	SMAPE	RRSE	RMSLE	RMSE
AR	3.416	0.258	0.268	1.207	0.342	4.321
NPAR	2.900	0.205	0.223	1.117	0.309	3.999
ARMA	2.971	0.217	0.228	1.122	0.310	4.018
NNA	3.147	0.212	0.243	1.213	0.327	4.343
NAÏVE	4.043	0.309	0.320	1.415	0.402	5.065
SNAÏVE	3.826	0.290	0.287	1.404	0.370	5.027

baseline models. This indicates that short-term rather than long-term values significantly affect the O_3 concentration. In addition to the above, the one-day predictive error was minimal compared to the other three horizons. Taking the NPAR model as an example, the MAE, MAPE, SMAPE, RRAE, RMSLE, and RMSE of the one-day prediction were 2.692, 0.192, 0.194, 1.022, 0.229, and 3.685, less than 2.879, 0.216, 0.217, 1.116, 0.309, and 3.990 for the two-day prediction; 2.944, 0.218, 0.217, 1.185, 0.316, and 3.987 for the three-day prediction; and 2.900, 0.205, 0.223, 1.117, 0.309, and 3.999 for the seven-day prediction. The horizon predictive effects of the AR, ARMA, and NNA models were the same as the NPAR model, indicating that the shorter the predictive horizon of the model, the better the predictive effect. The longer the predictive horizon, the worse the predictive effect. Thus, it can be seen from these results that the predictive error of the NPAR model was the smallest, and the predictive effect was the best compared to the rest within the proposed forecasting models and the baseline models. In addition, this also indicated that recent information was more effective in forecasting O_3 than old information after comparing the predictive errors of the four horizons of the three models.

Table 2. The O_3 in Campo de Marte Station: The mean forecast error for all models for a twenty-four-hour ahead out-of-sample forecast.

One-day-ahead (24 hours ahead)						
MODEL	MAE	SMAPE	MAPE	RRSE	RMSLE	RMSE
AR	4.6696	0.1171	0.1303	0.5597	0.1069	6.9269
NPAR	4.752	0.1182	0.1343	0.6667	0.1993	7.891
ARMA	5.2546	0.1274	0.1463	0.6295	0.223	7.6542
NNA	5.4503	0.131	0.1382	0.645	0.2066	7.8429
NAÏVE	5.1361	0.1245	0.1322	0.6035	0.1985	7.3385
SNAÏVE	10.4678	0.2587	0.4033	1.3213	0.5135	16.0659
Two-day-ahead (48 hours ahead)						
MODEL	MAE	MAPE	SMAPE	RRSE	RMSLE	RMSE
AR	5.909	0.198	0.158	0.732	0.301	8.905
NPAR	6.009	0.201	0.166	0.714	0.306	9.134
ARMA	6.485	0.213	0.169	0.787	0.315	9.818
NNA	6.688	0.205	0.175	0.788	0.307	9.827
NAÏVE	6.737	0.203	0.176	0.777	0.303	9.691
SNAÏVE	10.313	0.397	0.256	1.275	0.509	15.907
Three-day-ahead (72 hours ahead)						
MODEL	MAE	MAPE	SMAPE	RRSE	RMSLE	RMSE
AR	6.489	0.220	0.178	0.795	0.352	10.333
NPAR	6.695	0.231	0.184	0.808	0.358	10.491
ARMA	6.485	0.220	0.178	0.793	0.352	10.304
NNA	7.794	0.244	0.210	0.877	0.372	11.393
NAÏVE	7.064	0.228	0.191	0.833	0.362	10.821
SNAÏVE	10.011	0.367	0.264	1.178	0.510	15.309
Seven-day-ahead (168 hours ahead)						
MODEL	MAE	MAPE	SMAPE	RRSE	RMSLE	RMSE
AR	7.171	0.294	0.201	0.801	0.398	11.016
NPAR	7.288	0.295	0.204	0.801	0.399	11.021
ARMA	7.169	0.285	0.201	0.796	0.395	10.943
NNA	9.913	0.368	0.269	1.103	0.509	15.171
NAÏVE	10.963	0.356	0.294	1.061	0.458	14.588
SNAÏVE	10.156	0.329	0.274	0.973	0.434	13.384

In the same way, Table 2 contains the findings for the Campo de Marte Station. The accuracy mean errors of the six models are listed in this table for four horizons, such as one day, two days, three days, and seven days ahead: four from within the proposed component-based forecasting technique: the AR, NPAR, ARMA, and NNA models; and two baseline models: the naive and seasonal naive models. The following two inferences were derived from Table 2. The AR and ARMA models had very low mean accuracy errors. As shown in Table 2, the AR model had the best prediction ability, with accuracy mean errors (MAEs, MAPEs, SMAPEs, RRSE, RMSLE, and RMSE) of the one-day (4.670, 0.117, 0.130, 0.560, 0.107, and 6.927), two-day (5.909, 0.198, 0.158, 0.732, 0.301, and 8.905). However, mean errors are the best results shown by the ARMA models in terms of three-day and seven-day forecast accuracy, with the following metrics: three-day (6.485, 0.220, 0.178, 0.793, 0.352, and 10.304) and seven-day (7.169, 0.285, 0.201, 0.796, 0.395, and 10.943) ahead forecasts, respectively, which are less than the NPAR and NNA models within the proposed forecasting methodology and also significantly less than the baseline models. The NNA model had the weakest predictive effect and had substantially greater mean errors than the NPAR model. Despite this, the NPAR model produces the third-best results. When comparing the best model within the proposed forecasting technique to the baseline models (naive and season-naive models), Table 2 shows that the AR model outperforms the baseline models in one- and two-day forecasts, while the ARMA model outperforms the baseline models in three- and seven-day forecasts. Therefore, again, research demonstrated that short-term values have a greater impact on O_3 concentration than long-term values. Along with the aforementioned, the one-day forecast inaccuracy was modest compared to the other three timeframes. Taking the AR model as an example, the MAE, MAPE, SMAPE, RRAE, RMSLE, and RMSE of the one-day prediction were 4.670, 0.117, 0.130, 0.560, 0.107, and 6.927, less than 5.909, 0.198, 0.158, 0.732, 0.301, and 8.905 for the two-day prediction; 6.489, 0.220, 0.178, 0.795, 0.352, and 10.333 for the three-day prediction; and

7.171, 0.294, 0.201, 0.801, 0.398, and 11.016 for the seven-day prediction. The AR, ARMA, and NNA models all had the same horizon predictive results as the NPAR model, suggesting that the shorter the prediction horizon of the model, the better the predictive impact. The longer the predicted horizon, the less accurate the prediction. Thus, the prediction error of the AR and ARMA models was the least, and the predictive impact was the best when compared to the rest of the suggested forecasting models and baseline models. Furthermore, comparing the predicted errors of the four horizons of the three models revealed that recent information was more helpful in estimating ozone levels than ancient information.

Table 3. The O_3 in San Borja Station: The mean forecast error for all models for a twenty-four-hour ahead out-of-sample forecast.

One-day-ahead (24 hours ahead)						
MODEL	MAE	SMAPE	MAPE	RRSE	RMSLE	RMSE
AR	3.3365	0.2689	0.3442	0.6845	0.3543	4.1535
NPAR	3.1904	0.2505	0.2997	0.6502	0.3517	3.9458
ARMA	3.2937	0.2658	0.3273	0.6808	0.3553	4.1312
NNA	3.2018	0.2601	0.3643	0.653	0.3732	3.9624
NAÏVE	3.3159	0.2635	0.303	0.6809	0.3526	4.1318
SNAÏVE	5.8544	0.4466	0.78	1.1846	0.6638	7.1885
Two-day-ahead (48 hours ahead)						
MODEL	MAE	MAPE	SMAPE	RRSE	RMSLE	RMSE
AR	3.987	0.432	0.313	0.827	0.412	4.969
NPAR	3.758	0.440	0.294	0.765	0.404	4.596
ARMA	3.879	0.417	0.303	0.800	0.395	4.807
NNA	3.708	0.381	0.288	0.760	0.358	4.569
NAÏVE	4.007	0.389	0.311	0.829	0.410	4.980
SNAÏVE	5.781	0.767	0.440	1.185	0.657	7.122
Three-day-ahead (72 hours ahead)						
MODEL	MAE	MAPE	SMAPE	RRSE	RMSLE	RMSE
AR	4.000	0.486	0.316	0.861	0.466	5.167
NPAR	3.920	0.397	0.299	0.824	0.442	4.950
ARMA	3.923	0.399	0.305	0.836	0.453	5.022
NNA	3.961	0.521	0.308	0.831	0.473	4.988
NAÏVE	4.043	0.432	0.316	0.874	0.458	5.249
SNAÏVE	5.875	0.767	0.453	1.213	0.692	7.284
Seven-day-ahead (168 hours ahead)						
MODEL	MAE	MAPE	SMAPE	RRSE	RMSLE	RMSE
AR	4.975	0.468	0.404	1.054	0.552	6.276
NPAR	4.578	0.625	0.350	0.940	0.517	5.599
ARMA	5.020	0.594	0.393	1.043	0.514	6.207
NNA	4.491	0.442	0.341	0.928	0.505	5.527
NAÏVE	5.151	0.598	0.400	1.068	0.545	6.356
SNAÏVE	5.773	0.753	0.444	1.211	0.685	7.211

Likewise, preliminary results for the San Borja Station are presented in Table 3. The accuracy mean errors of the six models at four horizons, such as one day, two days, three days, and seven days ahead, are presented in Table 3: four from within the proposed component-based forecasting technique: the AR, NPAR, ARMA, and NNA models; and two baseline models: the naive and seasonal naive models. Table 3 yielded the following two findings. The mean accuracy errors of the NPAR and NNA models are minimal. As listed in Table 3, the NPAR model had the best forecasting effect for one-day and three-day ahead forecasting, with accuracy mean errors (MAEs, MAPEs, SMAPEs, RRSE, RMSLE, and RMSE) of the one-day ahead forecast (3.190, 0.251, 0.300, 0.650, 0.352, and 3.946), and three-day ahead forecasts (3.920, 0.397, 0.299, 0.824, 0.442, and 4.950), while the NNA model had the best forecasting effect for two-day and seven-day ahead forecasting, with mean errors of the two-day (3.708, 0.381, 0.288, 0.760, 0.358, and 4.569) and seven-day (4.491, 0.442, 0.341, 0.928, 0.505, and 5.527) ahead forecasts, respectively, less than the AR, and the ARMA models within the proposed forecasting methodology, and also significantly minimal to the baseline models. The AR model had the poorest predictive effect, with substantially greater mean errors than the ARMA model. Although the NPAR model produces the second-best performance, comparing the best model within

the proposed forecasting technique to the baseline models (naive and season-naive models), Table 5 shows that the NPAR model outperforms the baseline models. This suggests that short-term values had a more significant impact on O_3 concentration than long-term values. The one-day forecast error was also small compared to the other three forecasting timeframes. Taking the NPAR model as an example, the MAE, MAPE, SMAPE, RRAE, RMSLE, and RMSE of the one-day prediction were 3.190, 0.251, 0.300, 0.650, 0.352, and 3.946, less than 3.758, 0.440, 0.294, 0.765, 0.404, and 4.596 for the two-day prediction; 3.920, 0.397, 0.299, 0.824, 0.442, and 4.950 for the three-day prediction; and 4.578, 0.625, 0.350, 0.940, 0.517, and 5.599 for the seven-day prediction. The AR, ARMA, and NNA models all had the same horizon predictive results as the NPAR model, suggesting that the shorter the prediction horizon of the model, the better the predictive impact. The longer the forecast's horizon, the less accurate the forecast seems to be. Thus, it can be observed from these data that the NPAR model had the minimum predicted error and the best predictive impact when compared to the other proposed forecasting models and baseline models. Furthermore, comparing the forecasted errors of the four horizons of the three models revealed that recent information was more helpful in estimating ozone levels than historical information.

Table 4. The O_3 in Santa Anita Station: The mean forecast error for all models for a twenty-four-hour ahead out-of-sample forecast.

One-day-ahead (24 hours ahead)						
MODEL	MAE	SMAPE	MAPE	RRSE	RMSLE	RMSE
AR	3.7604	0.361	0.4296	1.0933	0.4317	4.8461
NPAR	3.1036	0.296	0.3415	0.955	0.3666	4.2332
ARMA	3.2337	0.3096	0.378	0.9786	0.3856	4.3378
NNA	3.1828	0.3064	0.3529	0.9679	0.3765	4.2901
NAÏVE	3.925	0.3692	0.4319	1.1339	0.4408	5.0262
SNAÏVE	5.2327	0.4827	0.6442	1.4868	0.6032	6.5903
Two-day-ahead (48 hours ahead)						
MODEL	MAE	MAPE	SMAPE	RRSE	RMSLE	RMSE
AR	3.890	0.446	0.369	1.131	0.447	5.054
NPAR	3.235	0.344	0.305	0.956	0.363	4.269
ARMA	3.313	0.384	0.314	0.977	0.387	4.366
NNA	3.252	0.349	0.308	0.964	0.371	4.306
NAÏVE	4.234	0.474	0.389	1.237	0.471	5.526
SNAÏVE	5.199	0.635	0.478	1.466	0.598	6.549
Three-day-ahead (72 hours ahead)						
MODEL	MAE	MAPE	SMAPE	RRSE	RMSLE	RMSE
AR	4.162	0.488	0.405	1.166	0.492	5.243
NPAR	3.584	0.385	0.349	1.022	0.415	4.595
ARMA	3.597	0.427	0.348	1.009	0.425	4.538
NNA	3.522	0.384	0.344	1.003	0.411	4.511
NAÏVE	4.484	0.510	0.418	1.260	0.505	5.666
SNAÏVE	5.195	0.636	0.486	1.454	0.610	6.537
Seven-day-ahead (168 hours ahead)						
MODEL	MAE	MAPE	SMAPE	RRSE	RMSLE	RMSE
AR	4.694	0.533	0.443	1.326	0.534	5.905
NPAR	3.565	0.373	0.346	1.042	0.413	4.640
ARMA	3.705	0.445	0.354	1.071	0.441	4.768
NNA	3.577	0.381	0.348	1.052	0.420	4.683
NAÏVE	5.472	0.625	0.478	1.532	0.574	6.819
SNAÏVE	5.129	0.626	0.480	1.456	0.604	6.479

Conversely, Table 4 tabulated the outcomes for the Santa Anita Station. This table illustrates the accuracy mean errors of the four horizons, such as one day, two days, three days, and seven days ahead, of the six models: four from within the proposed component-based forecasting technique: the AR, NPAR, ARMA, and NNA models; and two baseline models: the naive and seasonal naive models. The following two conclusions were drawn from Table 4. The mean accuracy errors of the NPAR model

were minimal. As shown in Table 4, the NPAR model had the best forecasting effect, with accuracy mean errors (MAEs, MAPEs, SMAPEs, RRSE, RMSLE, and RMSE) of the one-day (3.104, 0.296, 0.342, 0.955, 0.367, and 4.233), two-day (3.235, 0.344, 0.305, 0.956, 0.363, and 4.269), and seven-day (3.565, 0.373, 0.346, 1.042, 0.413, and 4.640) ahead forecasts, respectively, less than the AR, ARMA, and NNA models within the proposed forecasting methodology, and also significantly minimal to the baseline models (the naive and seasonal naive models). The predictive effect of the NNA model was the worst and was much higher than the mean errors of the AR and ARMA models. However, only in the case of three-day forecast accuracy, mean errors are the best results shown by the NNA models with the following metrics: MAE = 3.522, MAPE = 0.384, SMAPE = 0.344, RRSE = 1.003, RMSLE = 0.411, and RMSE = 4.511. Although the NPAR model also shows the second-best results, on the other hand, comparing the best model within the proposed forecasting approach with the baseline models (naive and season-naive models), it is confirmed from Table 3 that the NPAR model outperforms the baseline models. This indicated that short-term rather than long-term values significantly affect the O_3 concentration. Along with the aforementioned, the one-day forecast error was small compared to the other three spans. The MAE, MAPE, SMAPE, RRAE, RMSLE, and RMSE of the one-day prediction for the AR model were 3.7604, 0.361, 0.4296, 1.0933, 0.4317, and 4.8461, which were less than 3.890, 0.446, 0.369, 1.131, 0.447, and 5.054 for the two-day prediction; and 4.162, 0.488, 0.405, 1.166, 0.492, and 5.243 for the three-day prediction. The NPAR, ARMA, and NNA models all had the same horizon predictive effects as the AR model, demonstrating that the shorter the prediction horizon of the model, the better the predictive impact. The longer the forecasting horizon, the less accurate the forecast is. Thus, it can be observed from these data that the NPAR model had the minimum predicted error and the best predictive impact when compared to the other proposed forecasting models and baseline models. Furthermore, comparing the predicted errors of the four horizons of the three models revealed that recent information was more helpful in estimating ozone levels than past information.

In order to confirm the dominance of the best models for all monitoring stations listed in Tables 1-4, in this work, we performed the DM test on each pair of models. The null hypothesis is that the two models on the columns and rows are equally accurate, and the alternative hypothesis is that the model on the columns is more accurate than the model on the rows (using the loss-squared function). The results (DM-statistic) of the DM test are given in Table 5 (the Ate station), Table 6 (the CDM station), 7 (the SB station), and Table 8 (the STA station) of Metropolitan Lima. Thus, if the DM statistic is negative in these tables, the first predictive model (the column predictive model) is statistically better than the second predictive model (the row predictive model). Hence, the results of the Ate station show that the final super best (NPAR) model within all four best models and the considered two baseline models is statistically superior at the 5% significance level at all four forecasting horizons. However, in the CDM, the SB, and the STA stations, the final best models (the AR at one- and two-day ahead horizons and the ARMA at three- and seven-day ahead horizons), (the NPAR at one- and three-day ahead horizons and the ARMA at two- and seven-day ahead horizons) and (the NPAR at one-, three-, and seven-day ahead horizons and the NNA at two-day ahead horizon) are statistically superior to the other all considered models at the 5% level of significance.

Once the proposed component-based modeling and forecasting technique performance has been evaluated by accuracy performance measures (MAE, MAPE, SMAPE, RMSE, RMSLE, and RRSE) and a statistical test (the DM test), we then process the models for graphic analysis. For instance, we draw the scatter plots for each station using their respective best model obtained by accuracy mean

errors and a previous statistical test. Figure 2 displays the scatter plots for all considered monitoring stations, including Figure 2 (a) for the Ate station, Figure 2 (b) for the CDM station, Figure 2 (c) for the SB station, and Figure 2 (d) for the STA station. These figures show that the best models produce greater Pearson correlation coefficient values, which indicates that the correlation between forecast and actual O_3 values is highly significant. On the other hand, the forecasted and observed values for the supermodel in each monitoring station are plotted in Figure 3. In Figure 3, (a) for the Ate station, (b) for the CDM station, (c) for the SB station, and (d) for the STA station, forecasts of the best models follow the observed concentration of ozone very closely; to this, we can conclude that the best models in each considered station have accurate and efficient forecasts. Thus, from the descriptive statistical analysis, tests, and graphical results, we can be point that the proposed component-based modeling and forecasting technique is highly efficient and accurate in forecasting hourly O_3. In addition, within the proposed forecasting methodology, there are two classes of forecasting models: linear (the AR and the ARMA) and nonlinear (the NAPR and the NNA) time series models. As we confirm from the above results, the nonlinear models dominate overall, while in a few cases, the linear model outperforms the nonlinear models.

Table 5. The Diebold and Marino results for the Ate station: The DM-statistic values for all models are given in Table 1.

One-day-ahead (24 hours ahead)						
Models	AR	NPAR	ARMA	NNA	NAÏVE	SNAÏVE
AR	0.0000	-9.0887	-9.0935	-8.8935	-9.0899	5.6142
NPAR	9.0887	0.0000	2.8016	7.8766	7.0433	7.9687
ARMA	9.0935	-2.8016	0.0000	7.8831	6.9920	7.9692
NNA	8.8935	-7.8766	-7.8831	0.0000	-7.8895	7.9761
NAÏVE	9.0899	-7.0433	-6.9920	7.8895	0.0000	7.9698
SNAÏVE	-5.6142	-7.9687	-7.9692	-7.9761	-7.9698	0.0000
Two-day-ahead (48 hours ahead)						
Models	AR	NPAR	ARMA	NNA	NAÏVE	SNAÏVE
AR	0.0000	-9.5751	-9.5682	-9.4378	-9.5679	5.3401
NPAR	9.5751	0.0000	-3.0507	7.6242	7.0475	7.7160
ARMA	9.5682	3.0507	0.0000	7.6015	6.5041	7.7141
NNA	9.4378	-7.6242	-7.6015	0.0000	-7.6183	7.7234
NAÏVE	9.5679	-7.0475	-6.5041	7.6183	0.0000	7.7156
SNAÏVE	-5.3401	-7.7160	-7.7141	-7.7234	-7.7156	0.0000
Three-day-ahead (72 hours						
Models	AR	NPAR	ARMA	NNA	NAÏVE	SNAÏVE
AR	0.0000	-8.7263	-8.7151	-9.0370	-8.6979	4.5748
NPAR	8.7263	0.0000	-2.0191	6.1792	18.9793	6.2607
ARMA	8.7151	2.0191	0.0000	6.1427	11.0496	6.2578
NNA	9.0370	-6.1792	-6.1427	0.0000	-6.0526	6.2674
NAÏVE	8.6979	-18.9793	-11.0496	6.0526	0.0000	6.2511
SNAÏVE	-4.5748	-6.2607	-6.2578	-6.2674	-6.2511	0.0000
Seven-day-ahead (168 hours ahead)						
Models	AR	NPAR	ARMA	NNA	NAÏVE	SNAÏVE
AR	0.0000	-8.8059	-8.7952	-8.4630	-8.7729	14.7468
NPAR	8.8059	0.0000	2.9659	11.6282	19.9019	11.3933
ARMA	8.7952	-2.9659	0.0000	11.5031	19.9553	11.3841
NNA	8.4630	-11.6283	-11.5031	0.0000	-11.3278	11.3743
NAÏVE	8.7729	-19.9019	-19.9553	11.3278	0.0000	11.3712
SNAÏVE	-14.7468	-11.3933	-11.3841	-11.3743	-11.3712	0.0000

Table 6. Diebold and Marino results for the Campo de Marte station: The DM-statistic values for all models are given in Table 2.

One-day-ahead (24 hours ahead)						
Models	AR	NPAR	ARMA	NNA	NAÏVE	SNAÏVE
AR	0.0000	-2.8199	-2.8217	-2.8216	-2.8222	2.5844
NPAR	2.8199	0.0000	1.2125	2.6957	1.7028	2.6888
ARMA	2.8217	-1.2125	0.0000	2.8074	2.1681	2.6894
NNA	2.8216	-2.6957	-2.8074	0.0000	-2.8480	2.6887
NAÏVE	2.8222	-1.7028	-2.1681	2.8480	0.0000	2.6896
SNAÏVE	-2.5844	-2.6888	-2.6894	-2.6887	-2.6896	0.0000
Two-day-ahead (48 hours ahead)						
Models	AR	NPAR	ARMA	NNA	NAÏVE	SNAÏVE
AR	0.0000	-4.6521	-4.6491	-4.6552	-4.6501	4.0207
NPAR	4.6521	0.0000	-1.9907	4.4231	7.3595	4.2838
ARMA	4.6491	1.9907	0.0000	4.2369	4.0521	4.2827
NNA	4.6552	-4.4231	-4.2369	0.0000	-4.2581	4.2830
NAÏVE	4.6501	-7.3595	-4.0521	4.2581	0.0000	4.2828
SNAÏVE	-4.0207	-4.2838	-4.2827	-4.2830	-4.2828	0.0000
Three-day-ahead (72 hours ahead)						
Models	AR	NPAR	ARMA	NNA	NAÏVE	SNAÏVE
AR	0.0000	-6.1758	-6.1735	-6.1677	-6.1719	6.8057
NPAR	6.1758	0.0000	-3.0788	6.8135	10.4177	6.5771
ARMA	6.1735	3.0788	0.0000	6.5635	6.7891	6.5758
NNA	6.1677	-6.8135	-6.5635	0.0000	-6.5119	6.5758
NAÏVE	6.1719	-10.4177	-6.7891	6.5119	0.0000	6.5755
SNAÏVE	-6.8057	-6.5771	-6.5758	-6.5758	-6.5755	0.0000
Seven-day-ahead (168 hours ahead)						
Models	AR	NPAR	ARMA	NNA	NAÏVE	SNAÏVE
AR	0.00000	-6.26121	-6.26481	-6.24572	-6.25654	8.29033
NPAR	6.26121	0.00000	-10.89898	7.71715	10.43512	8.01439
ARMA	6.26481	10.89898	0.00000	8.04551	11.12267	8.01600
NNA	6.24572	-7.71715	-8.04551	0.00000	-7.32966	8.01583
NAÏVE	6.25654	-10.43512	-11.12267	7.32966	0.00000	8.01278
SNAÏVE	-8.29033	-8.01439	-8.01600	-8.01583	-8.01278	0.00000

Table 7. Diebold and Marino results for the San Borja station: The DM-statistic values for all models are given in Table 3.

One-day-ahead (24 hours ahead)						
Models	AR	NPAR	ARMA	NNA	NAÏVE	SNAÏVE
AR	0.0000	-3.7110	-3.7439	-3.7089	-3.7168	4.0280
NPAR	3.7110	0.0000	1.6320	3.7670	2.9423	3.8163
ARMA	3.7439	-1.6320	0.0000	5.4509	-0.3617	3.8394
NNA	3.7089	-3.7670	-5.4509	0.0000	-3.9880	3.8175
NAÏVE	3.7168	-2.9423	0.3617	3.9880	0.0000	3.8207
SNAÏVE	-4.0280	-3.8163	-3.8394	-3.8175	-3.8207	0.0000
Two-day-ahead (48 hours ahead)						
Models	AR	NPAR	ARMA	NNA	NAÏVE	SNAÏVE
AR	0.0000	-5.8812	-5.8247	-5.8246	-5.8486	5.9755
NPAR	5.8812	0.0000	-2.4739	8.4385	-1.4931	5.9179
ARMA	5.8247	2.4739	0.0000	5.8244	3.2554	5.8805
NNA	5.8246	-8.4385	-5.8244	0.0000	-6.6795	5.8819
NAÏVE	5.8486	1.4931	-3.2554	6.6795	0.0000	5.8965
SNAÏVE	-5.9755	-5.9179	-5.8805	-5.8819	-5.8965	0.0000
Three-day-ahead (72 hours ahead)						
Models	AR	NPAR	ARMA	NNA	NAÏVE	SNAÏVE
AR	0.0000	-6.0736	-6.0326	-6.0078	-6.0385	7.8289
NPAR	6.0736	0.0000	-3.1012	8.7842	-0.4728	6.6660
ARMA	6.0326	3.1012	0.0000	6.7054	5.2996	6.6326
NNA	6.0078	-8.7842	-6.7054	0.0000	-7.1164	6.6308
NAÏVE	6.0385	0.4728	-5.2996	7.1164	0.0000	6.6399
SNAÏVE	-7.8289	-6.6660	-6.6326	-6.6308	-6.6399	0.0000
Seven-day-ahead (168 hours ahead)						
Models	AR	NPAR	ARMA	NNA	NAÏVE	SNAÏVE
AR	0.0000	-13.1648	-13.0593	-13.0974	-13.1023	10.7537
NPAR	13.1648	0.0000	-4.0693	11.9360	-1.0432	12.2667
ARMA	13.0594	4.0693	0.0000	12.1186	6.6877	12.2212
NNA	13.0974	-11.9360	-12.1186	0.0000	-12.8686	12.2237
NAÏVE	13.1023	1.0432	-6.6877	12.8686	0.0000	12.2420
SNAÏVE	-10.7537	-12.2667	-12.2212	-12.2237	-12.2420	0.0000

Table 8. Diebold and Marino results for the Santa Anita station: The DM-statistic values for all models are given in Table 4.

One-day-ahead (24 hours ahead)						
Models	AR	NPAR	ARMA	NNA	NAÏVE	SNAÏVE
AR	0.0000	-5.2875	-5.3083	-5.2203	-5.2884	8.1294
NPAR	5.2875	0.0000	2.7675	6.2293	4.9491	6.1872
ARMA	5.3083	-2.7675	0.0000	6.7317	-0.0123	6.2071
NNA	5.2203	-6.2293	-6.7317	0.0000	-6.3213	6.1852
NAÏVE	5.2884	-4.9491	0.0123	6.3213	0.0000	6.1909
SNAÏVE	-8.1294	-6.1872	-6.2071	-6.1852	-6.1909	0.0000
Two-day-ahead (48 hours ahead)						
Models	AR	NPAR	ARMA	NNA	NAÏVE	SNAÏVE
AR	0.0000	-5.8369	-5.8138	-5.7600	-5.8183	7.8662
NPAR	5.8369	0.0000	-3.0381	7.0218	-0.2926	6.5382
ARMA	5.8138	3.0381	0.0000	6.5466	4.8688	6.5192
NNA	5.7600	-7.0218	-6.5466	0.0000	-6.6721	6.5179
NAÏVE	5.8183	0.2926	-4.8688	6.6721	0.0000	6.5245
SNAÏVE	-7.8662	-6.5382	-6.5192	-6.5179	-6.5245	0.0000
Three-day-ahead (72 hours ahead)						
Models	AR	NPAR	ARMA	NNA	NAÏVE	SNAÏVE
AR	0.0000	-7.3310	-7.2985	-7.2848	-7.3063	8.0393
NPAR	7.3310	0.0000	-3.4977	8.0225	0.2028	7.5900
ARMA	7.2985	3.4977	0.0000	7.4779	5.7668	7.5678
NNA	7.2848	-8.0225	-7.4779	0.0000	-7.6111	7.5720
NAÏVE	7.3063	-0.2028	-5.7668	7.6111	0.0000	7.5737
SNAÏVE	-8.0393	-7.5900	-7.5678	-7.5720	-7.5737	0.0000
Seven-day-ahead (168 hours ahead)						
Models	AR	NPAR	ARMA	NNA	NAÏVE	SNAÏVE
AR	0.0000	-6.3528	-6.3311	-6.2905	-6.3270	8.3365
NPAR	6.3528	0.0000	-3.4125	7.3201	-0.3216	6.9978
ARMA	6.3311	3.4125	0.0000	6.8884	6.4540	6.9795
NNA	6.2905	-7.3201	-6.8884	0.0000	-6.8547	6.9838
NAÏVE	6.3270	0.3216	-6.4540	6.8547	0.0000	6.9785
SNAÏVE	-8.3365	-6.9978	-6.9795	-6.9838	-6.9785	0.0000

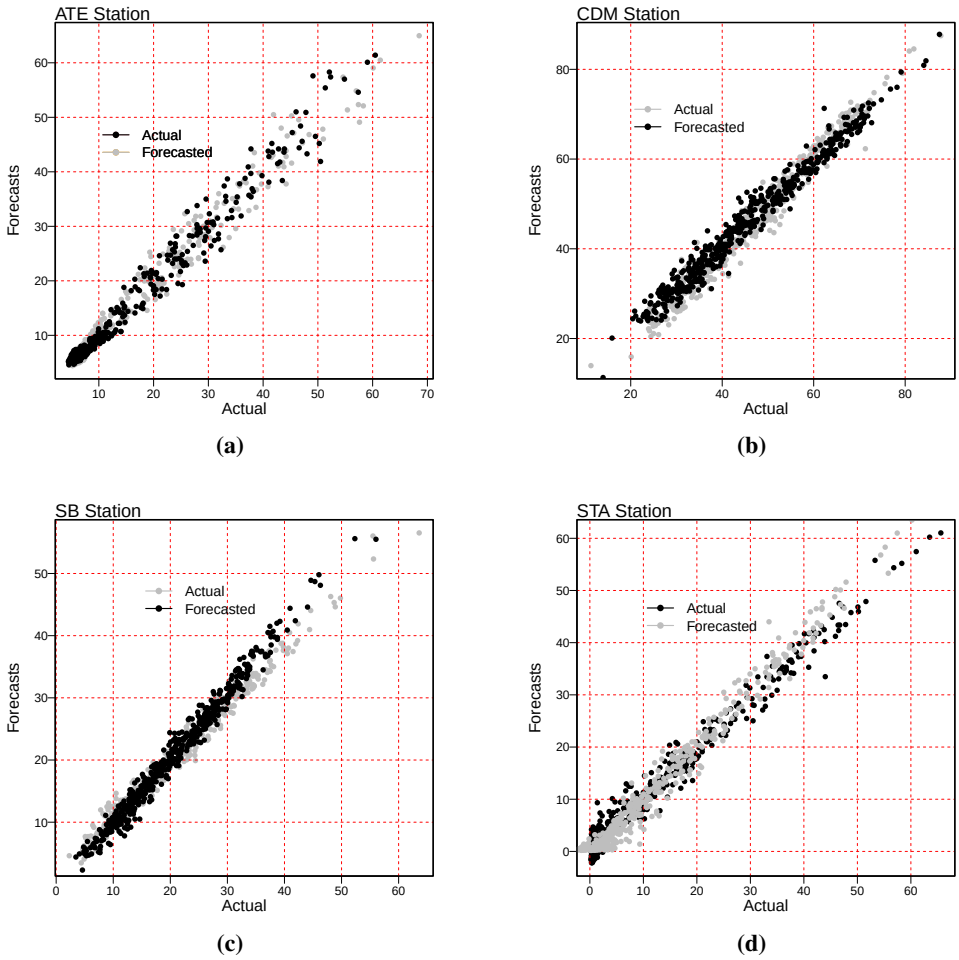

Figure 2. Correlation plot showing the O_3 for all stations. It displays the best forecasting models for the O_3, which are NPAR (located in the top-left), AR (top-right), ARMA (bottom-left), and NPAR (bottom-right).

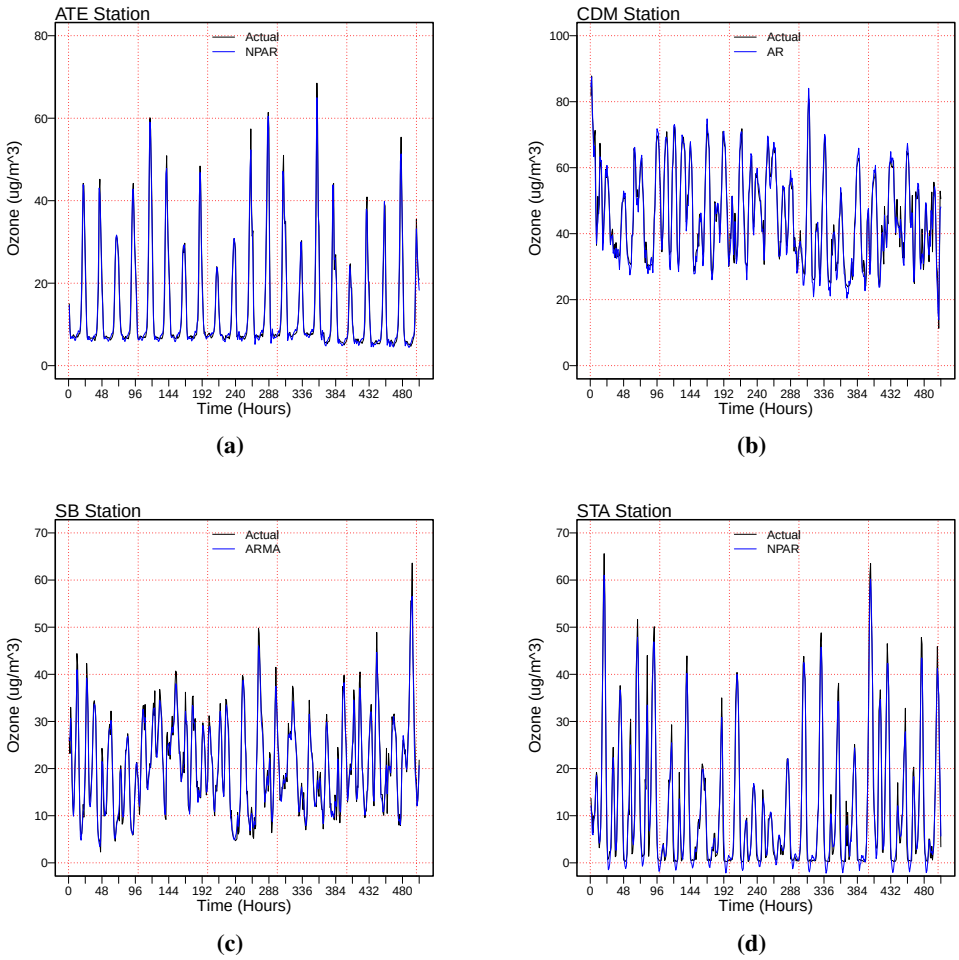

Figure 3. This graph displays the actual and forecasted data points for O_3 over 504 hours at four different stations: Ate Station (a), CDM Station (b), SB Station (c), and STA Station (d). The best model was used for each station.

In this sense, the ozone level is a prominent air pollutant in metropolitan areas, including four monitoring sites in Metropolitan Lima, Peru. When present in sufficient quantities, tropospheric ozone can have serious consequences for human health, including respiratory and cardiovascular disorders. Therefore, accurate and efficient forecasts short- to medium-term forecasts are more valuable for policymakers and decision-makers at the district and province levels. Thus, the authors recommended that the proposed component-based modeling and forecasting technique can be considered highly efficient and accurate in forecasting short- to medium-term hourly O_3.

4. Conclusions and future work directions

This research presents a component-based modeling and forecasting technique for predicting ozone levels in Metropolitan Lima, Peru. The method uses multiple linear regression, time series models, and data from four districts from 2017 to 2019. The hourly ozone time series is divided into deterministic and stochastic components, with four time series models used. The technique's performance was

validated using six standard accuracy measures, statistical tests, and graphical evaluations. Results showed that the nonparametric autoregressive model had the best forecasting effect for the Ate, CDM, SB, and STA stations. The neural network autoregressive model had the best forecasting effect for two-day and seven-day forecasting. In contrast, the nonparametric autoregressive model had the best impact on one-day, three-day, and seven-day forecasting. The technique demonstrated exceptional accuracy and efficiency in short- and medium-term forecasts of hourly O_3 levels in Lima, Peru.

However, the main limitation of this study is that it only presents hourly data on ozone levels. This could be improved by including additional external factors such as wind speed, temperature, wind direction, and humidity to enhance short-term predictions. Additionally, the study only used four district datasets in Lima, Peru. Still, it could be expanded to include other districts in Lima, different regions in Peru, and even globally to assess the effectiveness of the proposed component-based time series modeling and forecasting technique. Furthermore, the study employed only univariate time series models, which could be augmented by incorporating machine learning models like deep learning and artificial neural networks. These models could also be integrated into the current component-based time-series forecasting framework. Likewise, in other scenarios and with different data, for example, energy [48, 49], air pollution [50, 51, 52], and academic performance [53].

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

P.C. Rodrigues acknowledges financial support from the CNPq grant "bolsa de produtividade PQ-2" 309359/2022-8, Federal University of Bahia and CAPES-PRINT-UFBA, under the topic "Modelos Matemáticos, Estatísticos e Computacionais Aplicados às Ciências da Natureza.

Conflict of interest

The authors declare no conflict of interest.

References

1. Ghorani-Azam, A, Riahi-Zanjani B, Balali-Mood M (2016) Effects of air pollution on human health and practical measures for prevention in Iran. *J Res Med Sci* 21. https://doi.org/10.4103/1735-1995.189646
2. Hailstone J. Hailstone J (2023) https://www.forbes.com/sites/jamiehailstone/2023/03/07/nearly-nowhere-on-earth-safe-from-air-pollution-study-finds/?sh=1e0d9fd9da1d/ (accessed july 25, 2023).
3. Ordóñez C, Garrido-Perez J M, García-Herrera R (2020) Early spring near-surface ozone in Europe during the COVID-19 shutdown: Meteorological effects outweigh emission changes. *Sci Total Environ* 747: 141322. https://doi.org/10.1016/j.scitotenv.2020.141322

4. Mostafa M K, Gamal G, Wafiq A (2021) The impact of COVID 19 on air pollution levels and other environmental indicators-A case study of Egypt. *J Environ Manage* 277: 111496. https://doi.org/10.1016/j.jenvman.2020.111496
5. Gagliardi R V, Andenna C (2020) A machine learning approach to investigate the surface ozone behavior. *Atmosphere* 11: 1173. https://doi.org/10.3390/atmos11111173
6. Jaffe DA, Cooper OR, Fiore AM, et al. (2018) Scientific assessment of background ozone over the US: Implications for air quality management. *Elem Sci Anth* 6: 56. https://doi.org/10.1525/elementa.309
7. Lu H, Lyu X, Cheng H, et al.(2019) Overview on the spatial–temporal characteristics of the ozone formation regime in China. *Environmental Science: Processes & Impacts* 21: 916-929. https://doi.org/10.1039/C9EM00098D
8. Käffer M I, Domingos M, Lieske I, et al. (2019) Predicting ozone levels from climatic parameters and leaf traits of Bel-W3 tobacco variety. *Environ Pollut* 248: 471-477. https://doi.org/10.1016/j.envpol.2019.01.130
9. Li Y, Xue Y, Guang J, et al. (2018) Ground-level PM2. 5 concentration estimation from satellite data in the Beijing area using a specific particle swarm extinction mass conversion algorithm. *Remote Sens* 10: 1906. https://doi.org/10.3390/rs10121906
10. Velasco E, Retama A (2017) Ozone's threat hits back Mexico City. *Sustain Cities Soc* 31: 260-263. https://doi.org/10.1016/j.scs.2016.12.015
11. Carbo-Bustinza N, Belmonte M, Jimenez V, et al. (2022) A machine learning approach to analyse ozone concentration in metropolitan area of Lima, Peru. *Sci Rep* 12: 22084. https://doi.org/10.1038/s41598-022-26575-3
12. Cohen AJ, Brauer M, Burnett R, et al. (2017) Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *The Lancet* 389: 1907-1918. https://doi.org/10.1016/S0140-6736(17)30505-6
13. Iftikhar H, Khan M, Khan Z, et al. (2023). A Comparative Analysis of Machine Learning Models: A Case Study in Predicting Chronic Kidney Disease. *Sustainability* 15: 2754. https://doi.org/10.3390/su15032754
14. Jakovlev A R, Smyshlyaev S P et al. (2019) Numerical simulation of world ocean effects on temperature and ozone in the lower and middle atmosphere. *Russ Meteorol Hydrol* 44: 594-602. https://doi.org/10.3103/S1068373919090036
15. Gaudel A, Cooper OR, Ancellet G, et al. (2018) Assessment Report: Present-day distribution and trends of tropospheric ozone relevant to climate and global atmospheric chemistry model evaluation. *Elem Sci Anth* 6: 39. https://doi.org/10.1525/elementa.291
16. Rodríguez-Urrego D, Rodríguez-Urrego L (2020) Air quality during the COVID-19: PM2. 5 analysis in the 50 most polluted capital cities in the world. *Environ Pollut* 266: 115042. https://doi.org/10.1016/j.envpol.2020.115042
17. Rybarczyk Y, Zalakeviciute R (2018) Machine learning approaches for outdoor air quality modelling: A systematic review. *Appl Sci* 8: 2570. https://doi.org/10.3390/app8122570

18. Iftikhar H, Khan N, Raza MA, et al. (2024). Electricity theft detection in smart grid using machine learning. *Front Energy Res* 12: 1383090. https://doi.org/10.3389/fenrg.2024.1383090

19. Comrie A C (1997) Comparing neural networks and regression models for ozone forecasting. *J Air Waste Manage* 47: 653-663. https://doi.org/10.1080/10473289.1997.10463925

20. Carbo-Bustinza N, Iftikhar H, Belmonte M, et al. (2023). Short-term forecasting of Ozone concentration in metropolitan Lima using hybrid combinations of time series models. *Appl Sci* 13: 10514. https://doi.org/10.3390/app131810514

21. Harrou F, Fillatre L, Bobbia M, et al. (2013) Statistical detection of abnormal ozone measurements based on constrained generalized likelihood ratio test. In 52nd IEEE Conference on Decision and Control, Firenze, Italy, 10-13 December 2013. https://doi.org/10.1109/CDC.2013.6760673

22. Duenas C, Fernandez M C, Canete S, et al. (2005) Stochastic model to forecast ground-level ozone concentration at urban and rural areas. *Chemosphere* 61: 1379-1389. https://doi.org/10.1016/j.chemosphere.2005.04.079

23. Iftikhar H, Khan M, Turpo-Chaparro J E, et al. (2024). Forecasting stock prices using a novel filtering-combination technique: Application to the Pakistan stock exchange. *AIMS Math* 9: 3264-3288. https://doi.org/10.3934/math.2024159

24. Petetin H, Bowdalo D, Soret A, et al. (2020) Meteorology-normalized impact of the COVID-19 lockdown upon NO 2 pollution in Spain. *Atmos Chem Phys* 20: 19-11141. https://doi.org/10.5194/acp-20-11119-2020

25. Aljanabi M, Shkoukani M, Hijjawi M (2020) Ground-level ozone prediction using machine learning techniques: A case study in Amman, Jordan. *Int J Autom Comput* 17: 667-677. https://doi.org/10.1007/s11633-020-1233-4

26. Sousa S I V, Martins F G, Alvim-Ferraz M C, et al. (2007) Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environ Modell Softw* 22: 97-103. https://doi.org/10.1016/j.envsoft.2005.12.002

27. Chelani A B (2010) Prediction of daily maximum ground ozone concentration using support vector machine. *Environmental monitoring and assessment* 162: 169-176. https://doi.org/10.1007/s10661-009-0785-0

28. Ren X, Mi Z, Georgopoulos P G (2020) Comparison of Machine Learning and Land Use Regression for fine scale spatiotemporal estimation of ambient air pollution: Modeling ozone concentrations across the contiguous United States. *Environ Int* 142: 105827. https://doi.org/10.1016/j.envint.2020.105827

29. Yafouz A, AlDahoul N, Birima AH, et al. (2022) Comprehensive comparison of various machine learning algorithms for short-term ozone concentration prediction. *Alex Eng J* 61: 4607-4622. https://doi.org/10.1016/j.aej.2021.10.021

30. Pan Q, Harrou F, Sun Y A (2023) comparison of machine learning methods for ozone pollution prediction. *J Big Data* 10: 63. https://doi.org/10.1186/s40537-023-00748-x

31. Iftikhar H, Zafar A, Turpo-Chaparro J E, et al. (2023) Forecasting Day-Ahead Brent Crude Oil Prices Using Hybrid Combinations of Time Series Models. *Mathematics* 16: 3548. https://doi.org/10.3390/math11163548

32. Iftikhar H, Bibi N, Canas Rodrigues P, et al. (2023) Multiple Novel Decomposition Techniques for Time Series Forecasting: Application to Monthly Forecasting of Electricity Consumption in Pakistan. *Energies* 16: 2579. https://doi.org/10.3390/en16062579
33. Alshanbari H M, Iftikhar H, Khan F, et al. (2023). On the Implementation of the Artificial Neural Network Approach for Forecasting Different Healthcare Events. *Diagnostics* 13: 1310. https://doi.org/10.3390/diagnostics13071310
34. Iftikhar H (2018) Modeling and Forecasting Complex Time Series: A Case of Electricity Demand. Master's Thesis, Quaidi-Azam University, Islamabad, Pakistan, 1-94.
35. Shah I, Iftikhar H, Ali S (2020) Modeling and forecasting medium-term electricity consumption using component estimation technique. *Forecasting* 2: 163–179. https://doi.org/10.3390/forecast2020009
36. Iftikhar H, Turpo-Chaparro J E, Canas Rodrigues P, et al. (2023). Day-Ahead Electricity Demand Forecasting Using a Novel Decomposition Combination Method. *Energies* 16: 6675. https://doi.org/10.3390/en16186675
37. Shah I, Iftikhar H, Ali S, et al. (2019) Short-term electricity demand forecasting using components estimation technique. *Energies* 12: 2532. https://doi.org/10.3390/en12132532
38. Van Buuren S, Oudshoorn C G (2000) Multivariate imputation by chained equations.
39. Iftikhar H, Turpo-Chaparro J E, Canas Rodrigues P, et al. (2023). Forecasting Day-Ahead Electricity Prices for the Italian Electricity Market Using a New Decomposition—Combination Technique. *Energies* 16: 6669. https://doi.org/10.3390/en16186669
40. Diebold F X, Mariano R S (2022) Comparing predictive accuracy. *J Bus Econ Stat* 20: 134–144. https://doi.org/10.1198/073500102753410444
41. Iftikhar H, Khan M, Khan M S, et al. (2023). Short-Term Forecasting of Monkeypox Cases Using a Novel Filtering and Combining Technique. *Diagnostics* 13: 1923. https://doi.org/10.3390/diagnostics13111923
42. Shah I, Iftikhar H, Ali S (2022) Modeling and forecasting electricity demand and prices: A comparison of alternative approaches. *J Math* 2022: 3581037. https://doi.org/10.1155/2022/3581037
43. Iftikhar H, Daniyal M, Qureshi M, et al. (2023). A hybrid forecasting technique for infection and death from the mpox virus. *Digit Health* 9: 20552076231204748. https://doi.org/10.1177/20552076231204748
44. Dickey D A, Fuller W A (1979) Distribution of the estimators for autoregressive time series with a unit root. *J Am Stat Assoc* 74: 427–431. https://doi.org/10.1080/01621459.1979.10482531
45. Romero Y, Diaz C, Meldrum I, et al. (2020) Temporal and spatial analysis of traffic–Related pollutant under the influence of the seasonality and meteorological variables over an urban city in Peru. *Heliyon* https://doi.org/10.1016/j.heliyon.2020.e04029
46. Leon C A M, Felix M F M, Olivera C A C, et al (2022) Influence of Social Confinement by COVID-19 on Air Quality in the District of San 503 Juan de Lurigancho in Lima, Perù. *Chem Eng Trans* 91: 475–480.

47. Aaker D A, Jacobson R (1987). The sophistication of 'naive'modeling. *Int J Forecast* 3: 449-451. https://doi.org/10.1016/0169-2070(87)90039-2

48. Gonzales Javier L L, Calili Rodrigo F, Souza Reinaldo C, et al. (2016) Simulation of the energy efficiency auction prices in Brazil. *Renew Energ Power Qual J* 1: 574-579. https://doi.org/10.24084/repqj14.396

49. López-Gonzales J L, Souza RC, Da Silva FLC, et al. (2020) Simulation of the energy efficiency auction prices via the markov chain monte carlo method. *Energies* 13: 4544. https://doi.org/10.3390/en13174544

50. da Silva KLS, López-Gonzales J L, Turpo-Chaparro JE, et al. Spatio-temporal visualization and forecasting of PM10 in the Brazilian state of Minas Gerais. *Sci Rep* 13: 3269. https://doi.org/10.1038/s41598-023-30365-w

51. Jeldes N, Ibacache-Pulgar G, Marchant C, et al. (2022) Modeling Air Pollution Using Partially Varying Coefficient Models with Heavy Tails. *Mathematics* 10: 3677. https://doi.org/10.3390/math10193677

52. Cabello-Torres RJ, Estela MAP, Sánchez-Ccoyllo O, et al. (2022) Statistical modeling approach for PM10 prediction before and during confinement by COVID-19 in South Lima, Perú. *Sci Rep* 12: 1. https://doi.org/10.1038/s41598-022-20904-2

53. Orrego Granados D, Ugalde J, Salas R, et al. (2022) Visual-Predictive Data Analysis Approach for the Academic Performance of Students from a Peruvian University. *Appl Sci* 12: 11251. https://doi.org/10.3390/app122111251

AIMS Press