



Research article

Forecasting the concentration of NO₂ using statistical and machine learning methods: A case study in the UAEAishah Al Yammahi^{a,*}, Zeyar Aung^{a,b}^a Department of Electrical Engineering and Computer Science, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates^b Center for Catalysis and Separation (CeCaS), Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates

ARTICLE INFO

Keywords:

Machine learning
ARIMA
SARIMA
LSTM
NAR
Classical statistics
NO₂

ABSTRACT

Nitrogen dioxide (NO₂) is the most active pollutant gas emitted in the industrial era and is highly correlated with human activities. Tracking NO₂ emissions and predicting their concentrations represent important steps toward controlling pollution and setting rules to protect people's health indoors, such as in factories, and in outdoor environments. The concentration of NO₂ was affected by the COVID-19 lockdown period and decreased because of restrictions on outdoor activities. In this study, the concentration of NO₂ was predicted at 14 ground stations in the United Arab Emirates (UAE) during December 2020 based on training over a full time period of two years (2019–2020). Statistical and machine learning models, such as autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving average (SARIMA), long short-term memory (LSTM), and nonlinear autoregressive neural network (NAR-NN), are used with both open- and closed-loop architectures. The mean absolute percentage error (MAPE) was used to evaluate the performance of the models, and the results ranged from “very good” (MAPE of 8.64% at the Liwa station with the closed loop) to “acceptable” (MAPE of 42.45% at the Khadejah School station with the open loop). The results show that the predictions based on the open loop are generally better than those based on the closed loop because they yield statistically significantly lower MAPE values. For both loop types, we selected stations exhibiting the lowest, medium, and highest MAPE values as representative cases. In addition, we demonstrated that the MAPE value is highly correlated with the relative standard deviation of NO₂ concentration values.

1. Introduction

Atmospheric pollution has a number of impacts on human health, particularly in urban environments [1]. The concentration of pollutants corresponds to the population distribution among areas due to human activities [2–4]. One of the most important atmospheric components that has a direct relationship with pollution is nitrogen dioxide (NO₂), which is emitted mainly from diesel and petrol engines as reported in [5], with road transport contributing approximately 40% of the land-based NO_x emissions in European countries. Developed countries are facing numerous environmental and social issues related to air pollution. The establishment of ground monitoring stations to check air quality over the time is vital, as indicated by numerous studies all over the world such as those in USA [6], China [7], India [8], Brazil [9], Italy, and Sweden [10]. In addition, future air quality must be predicted to help governments formulate policies related to traffic control when the air is extremely polluted and outdoor activities [11].

* Corresponding author.

E-mail address: aishah.alyammahi@ku.ac.ae (A. Al Yammahi).

Time series modeling and forecasting are widely used in many research fields, such as economics, medicine, civil engineering, and climate [12–15]. Many models have been used for time series forecasting based on classical, machine learning, and deep learning methods [16,17]. Classical methods include the autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA) [18]. The ARIMA model has been used in many climate and environmental studies, such as forecasting temperature, precipitation, air pollution for two locations in India [19,20], as well as in estimating the wind speed, rainfall, air quality, and NO₂, NO, CO, and O₃ contents in many other geographical locations [21–26]. Many studies have used machine learning models, such as artificial neural networks (ANNs), support vector machines (SVMs), and deep learning prediction models, such as long short-term memory (LSTM), to predict the concentration of CO₂ [18,27–29].

A previous study [30] predicted air pollution based on pollutants (CO₂, CO, NO_x, SO₂, SO₃, and SPM) in Iran's atmosphere used the ARIMA, SVM, and TSVR models, a hybrid of the ARIMA and SVM models, and a hybrid of the ARIMA and TSVM models. Hybrid models with a combination of autoregressive parts are used to determine the minimum forecasting error. Data were collected from the statistical center in Iran between 1978 and 2016 for six pollutants (CO₂, CO, NO_x, SO₂, SO₃, and SPM), and the minimum error was obtained using the hybrid model (ARIMA-TSVM) among the other models. The section criteria included the root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). Fossil fuels have been reported to contribute the most to air pollution in Iran [30].

[31] focused only on the emissions of CO₂ in nine Asian countries (Japan, Bangladesh, China, Pakistan, India, Sri Lanka, Iran, Singapore, and Nepal) based on different sources of emissions, including heat and electricity, manufacturing industries, residential and commercial buildings, transport, and other sources. Every country presents different contributions from various industrial sectors. Annual data gathered from the World Bank from 1971 to 2014 and CO₂ emissions were predicted using ARIMA and simple exponential smoothing (SES) models. The minimum forecasting error was evaluated using the forecast mean absolute error (FMAE) [31]. The results showed that CO₂ emissions increased in Pakistan, Bangladesh, India, Iran, and Sri Lanka because of the use of heat and electricity. In China, the main causes of CO₂ emissions are residential buildings, commercial buildings, and transportation. However, specific sectors did not have a major effect on emissions in Nepal and Singapore. The best prediction model with the minimum errors for Japan, China, India, Iran, and Singapore was ARIMA, whereas the best model for forecasting CO₂ emissions in Pakistan and Sri Lanka was SES. Both models are equally appropriate for Nepal and Bangladesh [31].

World Bank data were also used to predict the annual CO₂ emissions (per capita) of the European Union for the period between 1960 and 2014 [21] using the ARIMA (1,1,1)-autoregressive conditional heteroscedasticity (ARCH) (1) model. The model parameters were estimated using the maximum-likelihood approach. The model was examined using dynamic (n-step ahead forecasts) and static (one step ahead) processes, although the results showed that the static procedure provided better predictions than the dynamic procedure. Despite the development of green energy plans, CO₂ emissions have been continuously increasing. The prediction models have presented different levels of accuracy in various areas and industrial sectors.

Similarly, the concentration of CH₄ was predicted by [32]. CH₄ measurements were performed using a Picarro G2401 cavity ring-down spectrometer in the Arctic island of Belyy, Russia, in the surface layer of the atmosphere. Hourly measurements collected over approximately a week (170 h) in the summer of 2015 were used to assess the forecasting model for methane concentration. The first 150 h were used as the training dataset, and the remaining 20 h were used for testing. Four models based on the artificial neural network (ANN), namely, nonlinear autoregressive neural network with an external input (NARX), Elman neural network (ENN), and multilayer perceptron (MLP) were used. The selection criteria were based on four minimum error calculations: MAE, RMSE, root mean squared relative error (RMSRE), and normalized root mean squared error (NRMSE). The best accuracy was achieved using NARX [32, 33]. The CH₄ concentration, temperature, humidity, and pressure values were determined at the same location (Belyy Island, Russia) for approximately two months (July and August 2017) in another study [34]. The data were averaged hourly, and 1175 data points were used. The first 1103 points were employed as the training data, whereas the remaining 72 time points were the testing data for time series forecasting. The accuracy selection criteria were the same as that in Ref. [32] except that NRMSE was excluded and Willmott's index of agreement and standard deviation was included.

The LSTM-RNN was used to predict one-day-ahead air quality based on the concentrations of CO, NO₂, O₃, PM₁₀, and SO₂, which were sampled at 22 sites in urban ground stations located in Madrid, Spain. The daily mean was computed from hourly data gathered over a period of 12 years (2001–2013) [35]. Similarly, NO₂ concentrations were predicted using ARIMAX (6,1,6) and CNN-LSTM in the study [36]. The results show that CNN-LSTM provides better results than ARIMAX with a mean squared error (MSE) of 0.21.

Another study [37] focused on developing a spatiotemporal model of the daily concentrations of three pollutants (PM_{2.5}, NO₂, and CO) across 50 sites in Pittsburgh, Pennsylvania, USA, between August 2016 and December 2017. The land used regression (LUR) and land use random forest (LURF) methods were examined and tested using the coefficient of determination (R²). The LURF models outperformed the LUR model based on an R² value of 0.10–0.19.

The Prophet forecasting model (PFM) was used to predict short- and long-term concentrations of 6 air pollutants PM_{2.5}, PM₁₀, O₃, NO₂, SO₂, and CO in Seoul, South Korea [38]. Hourly data were measured at 25 sites available on the Seoul Open Data Plaza for three years from 2017 to 2019. Seasonality, holidays, and errors were considered in the Prophet forecasting model (PEM). The performance of the model was evaluated using the MSE, MAE, and RMSE. The proposed PFM outperformed previous PEMs in the literature by increasing the number of pollutants from three to six and increasing the prediction time from one day to one year. This study encourages the use of additional data to achieve more accurate results.

In the study by [39], hourly measurements of NO, NO₂, NO_x, PM₁₀, SO₂ and ground-level O₃ were recorded in the town of Blagoevgrad, Bulgaria from an ExEA monitoring station. ARIMA and SARIMA time series analyses were performed to study a one-year period starting from September 1, 2011 and extending to August 31, 2012. The data were used to predict the measurements three days ahead. Factor analyses, such as principal component analysis (PCA), Promax rotation, multicollinearity between the six pollutants, and

Table 1
Summary of related works on air pollutant studies in various geographical areas.

Authors	Publication Year	Air Pollutants Studied	Geographical Region	Data Source	Period of Study	Time Resolution	Methods Used	Evaluation Criteria
Pakrooh & Pishbahar	2019	CO ₂ , CO, NO _x , SO ₂ , SO ₃ , and SPM	Iran	Statistical Center of Iran	1978 to 2016	yearly	ARIMA, SVM, TSVR, ARIMA-SVM, ARIMA-TSVR	MAE, MAPE, RMSE
Fatima	2019	CO ₂	Japan, Bangladesh, China, Pakistan, India, Sri Lanka, Iran, Singapore, and Nepal	World Bank data	1971 to 2014	yearly	ARIMA, Simple exponential smoothing (SES)	FMAE
Dritsaki & Dritsaki	2020	CO ₂	European Union	World Bank data	1960 to 2014	yearly	ARIMA (1,1,1)-ARCH(1)	dynamic (n-steps ahead), static (one step ahead)
Sergeev	2018	CH ₄	Arctic island Belyy, Russia	means of a cavity ring-down spectrometer Picarro G2401	~1 week (170 h) in 2015	hourly	ANN, NARX, ENN, MLP	RMSRE, NRMSRE
Navares & Aznarte	2020	CO, NO ₂ , O ₃ , PM ₁₀ , and SO ₂	Madrid, Spain	–	–	–	CNN-LSTM, ARIMAX	–
Jain	2021	PM _{2.5} , NO ₂ , and CO	Pittsburgh, Pennsylvania, USA	50 ground stations	Aug 2016 to Dec 2017	daily	Land used regression (LUR), Land use random forest (LURF)	R ²
Shen	2020	PM _{2.5} , PM ₁₀ , O ₃ , NO ₂ , SO ₂ , and CO	Seoul, South Korea	25 sites Seoul Open Data Plaza	2017 to 2019	hourly	Prophet forecasting model (PFM)	MSE, MAE, RMSE
Gocheva-Ilieva	2014	NO, NO ₂ , NO _x , PM ₁₀ , SO ₂ , and O ₃	Bulgaria	monitoring station of ExEA	Sep 2011 to Aug 2012	hourly	ARIMA, SARIMA	–

the Box-Jenkins method, were also applied.

The related works on the forecasting of air pollutants are summarized in Table 1.

With respect to the United Arab Emirates (UAE), previous studies that have applied various analysis methods for air pollutants, including NO₂, have used both ground station data and satellite data [40]. Among forecasting studies of air pollutants in the UAE, [41] conducted a high-level study on CO₂ forecasting in six Gulf countries, including the UAE, using the World Bank Database (WBD) with the ARIMA method.

To the best of our knowledge, previous studies have not performed predictive analyses (i.e., forecasting) of future atmospheric NO₂ concentrations in the UAE using data from a network of several ground stations. We believe that this study is the first to address this issue. We applied both the statistical and machine learning models to forecast the NO₂ concentration levels of the 14 ground stations in the Emirate of Abu Dhabi of the UAE with 1–31 day-ahead forecasts using both the open-loop and closed-loop setups. We achieved acceptable accuracies ranging from “very good” to “acceptable” among the 14 stations, which were measured in terms of the mean absolute percentage error (MAPE).

2. Data used

The United Arab Emirates (UAE) is located between 22.5° and 26.5° N latitude and 51.6°–56.5° E longitude. NO₂ measurement data for 14 sites in Abu Dhabi were provided by the Environment Agency of Abu Dhabi for two years (2019–2020). The 14 ground stations are distributed in the Abu Dhabi region, Al Dhafra region, and Al Ain region, as shown in Table 2 and the map of the UAE in Fig. 1. The ground station locations were classified into different categories: urban, suburban, and rural. The NO₂ concentrations measured at the ground stations were recorded hourly (i.e., 24 readings per day) in units of $\mu\text{g}/\text{m}^3$.

Some data values were missing, and they were replaced by linearly interpolated/extrapolated values. Outlier values were identified as those greater than or less than three standard deviations from the mean (i.e., covering 99.7% of the data) [42]. These outlier values were clipped (rounded down/up) to three standard deviations from the mean.

We illustrate the daily minimum, average, and maximum NO₂ concentration values of the three selected stations, one from each region, namely (i) Khadejah School station (urban, Abu Dhabi Capital Region), (ii) Ruwais station (suburban, Al Dhafra Region), and (iii) Al Quaa station (rural, Al Ain Region) for the period 2019–2020 in Fig. 2. We can observe from the graphs that the daily maximum NO₂ concentration values fluctuate significantly when compared to the daily average and minimum values at all three stations.

In this study, we focused on daily average NO₂ concentrations. The Pearson correlation coefficients [43] of the daily average NO₂ concentration values of the 14 stations during the period 2019–2020 are depicted as a heatmap in Fig. 3. Generally good correlations are observed among the stations in the Abu Dhabi Capital Region (except for the Baniyas School). The stations in Abu Dhabi were geographically close to each other. Three stations in the Al Ain region (excluding Al Quaa) also exhibited good correlations. The correlation values were relatively low for the stations in the Al Dhafra region (except for the Gayathi School and Ruwais).

3. Forecasting methods

3.1. Classical statistical methods

Time-series data are a sequence of numerical values in which each value is associated with a specific timestamp. Time-series data were classified as stationary or non-stationary. Patterns were not observed in the stationary data, while patterns, trends, and seasonality were observed in the non-stationary data due to the variations in the mean and variance over the time sequence. To apply statistical prediction models, such as the ARIMA and seasonal ARIMA (SARIMA), non-stationary data must be converted to stationary data. The differencing technique is used to convert nonstationary data into stationary data by calculating the difference between two consecutive observations. This differencing reduces the existence of trends and seasonality in the data. Such differencing is ordered, such as first- and second-order, as shown in Eqs. (1) and (2). If first-order differencing has no effect on creating stationary data, then

Table 2
Locations of the ground stations in UAE.

	Name	Location Category	Region
1	Hamdan Street	Urban Traffic	Abu Dhabi Capital Region
2	Khadejah School	Urban Background	Abu Dhabi Capital Region
3	Khalifa School	Suburban Background	Abu Dhabi Capital Region
4	Baniyas School	Suburban Background	Abu Dhabi Capital Region
5	Bain Al Jessrain	Suburban Background	Abu Dhabi Capital Region
6	Al Mafraq	Suburban Industrial	Abu Dhabi Capital Region
7	Bida Zayed	Suburban Background	Al Dhafra Region
8	Gayathi School	Suburban Background	Al Dhafra Region
9	Ruwais	Suburban Industrial	Al Dhafra Region
10	Liwa	Rural Background	Al Dhafra Region
11	Al Quaa	Rural Background	Al Ain Region
12	Sweihan	Suburban Background	Al Ain Region
13	Al Tawia	Suburban Background	Al Ain Region
14	Zakher	Urban Background	Al Ain Region

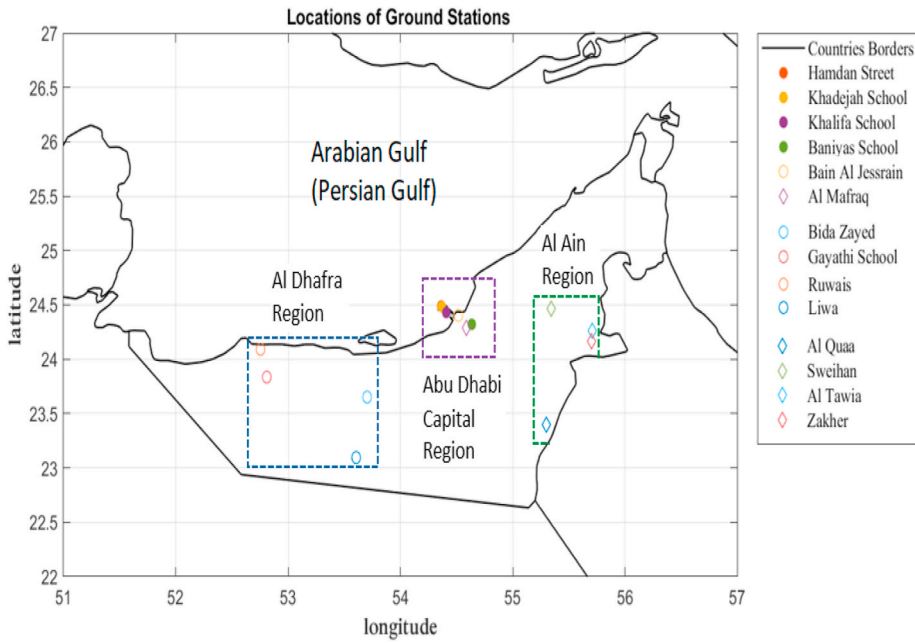


Fig. 1. Locations of 14 ground stations on the map of the UAE. (Note: Stations 1–3 are located relatively close to each other in Abu Dhabi City.) The X-axis represents in the longitude degrees (East) and the Y-axis represents the latitude degrees (North).

second-order differencing should be applied. The difference between the observation and seasonal data was computed within the same season. The first-order seasonal differencing is presented in Eq. (3). After removing the seasonality by differencing, the data were used for predicting the variable with respect to time, considering four assumptions about the data. First, anomalies or outliers were not observed in the data. Second, the data were univariate, and the regression was built based on its previous value. Third, the data were stationary with a constant mean and variance. Fourth, the parameters and error terms of the statistical prediction model are constant over time [44].

$$y_t' = y_t - y_{t-1} \tag{1}$$

$$y_t'' = y_t - 2y_{t-1} + y_{t-2} \tag{2}$$

$$y_s' = y_t - y_{t-m} \tag{3}$$

where.

- y_t' : time series after first-order differencing;
- y_t : nonstationary time series;
- y_{t-1} : observation at time step $t - 1$;
- y_t'' : time series after second-order differencing;
- y_{t-2} : observation at time step $t - 2$;
- y_s' : seasonal time series after first-order differencing;
- y_{t-m} : observation at time step $t - m$, where m is the number of time steps corresponding to a seasonal period.

3.1.1. Autoregressive integrated moving average (ARIMA (p,d,q))

ARIMA combines two regression models: autoregressive (AR) and moving average (MA). It has three parameters for the time steps: (p) for autoregression, (d) for differencing order, and (q) for moving average. The autoregression model predicts time series dependence based on its past value, whereas the moving average forecasts the time series dependence based on its previous predicted error to predict the variable for later time steps. The AR, MA, and ARIMA (which combines AR and MA in addition to differencing (d)) models are presented in Eqs. (4)–(6), respectively. The general representation of the ARIMA model is given in Eq. (7) [44,45]. If the data are stationary, then ARMA is applicable, which is a combination of AR and MA [46].

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \dots + \phi_p y_{t-p} + \epsilon_t \tag{4}$$

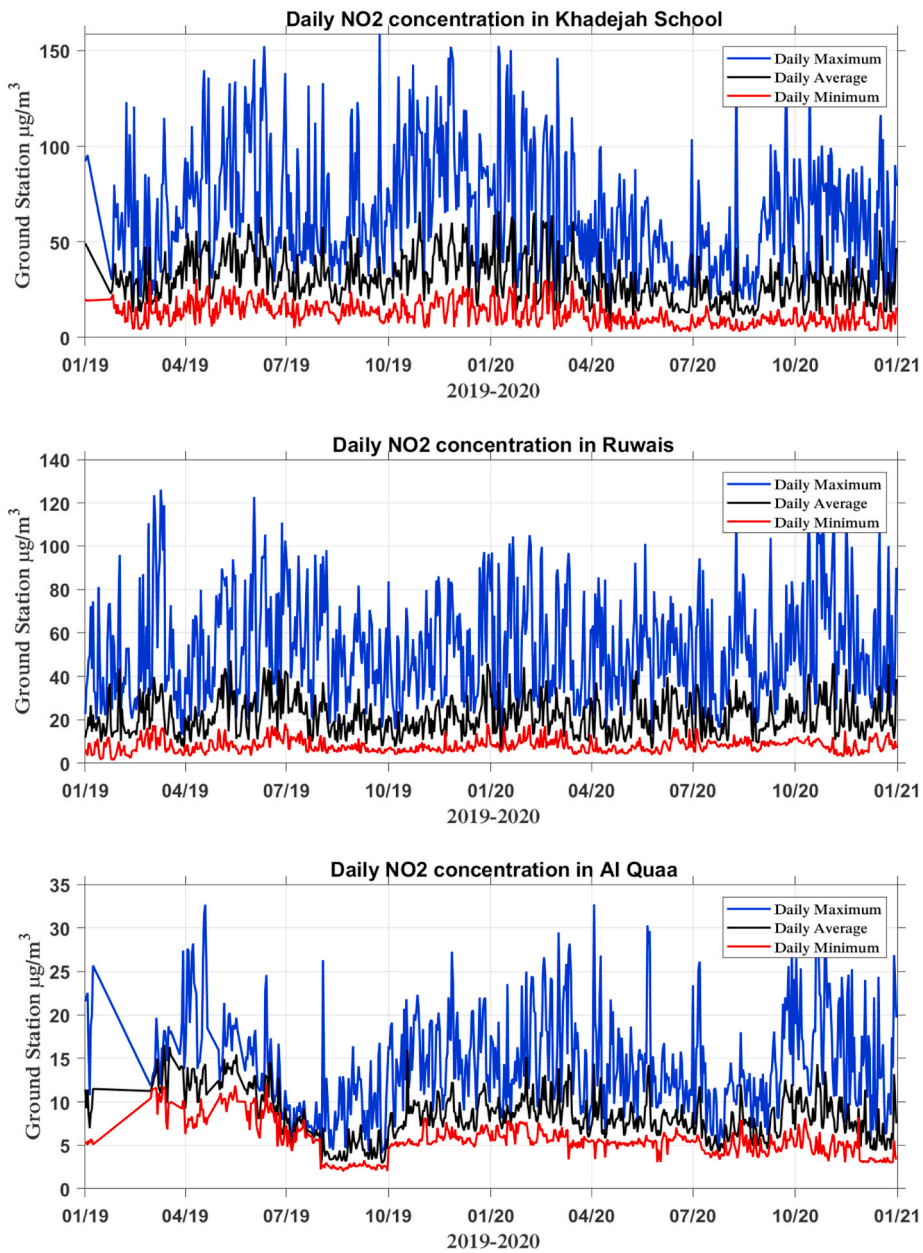


Fig. 2. Daily NO₂ concentration values in the three selected stations: Khadejah School station (urban, Abu Dhabi Capital Region), Ruwais station (suburban, Al Dhafra Region), and Al Quaa station (rural, Al Ain Region) in 2019–2020. (Note: the straight lines are the missing values replaced by linearly interpolated/extrapolated values.)

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \dots + \theta_q \varepsilon_{t-q} \tag{5}$$

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{6}$$

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \tag{7}$$

3.1.2. Seasonal autoregressive integrated moving average (SARIMA (p,d,q) (P,D,Q)_m)

SARIMA has an additional seasonal term compared with the non-seasonal ARIMA model, and this term is (P,D,Q)_m, which corresponds to seasonality in m time steps in a specific seasonal period. The SARIMA model is expressed by Eq. (8). B is the backshift operator expressed by Eq. (9). The expressions for non-seasonal and seasonal autoregression (AR), moving average (MA), seasonal AR

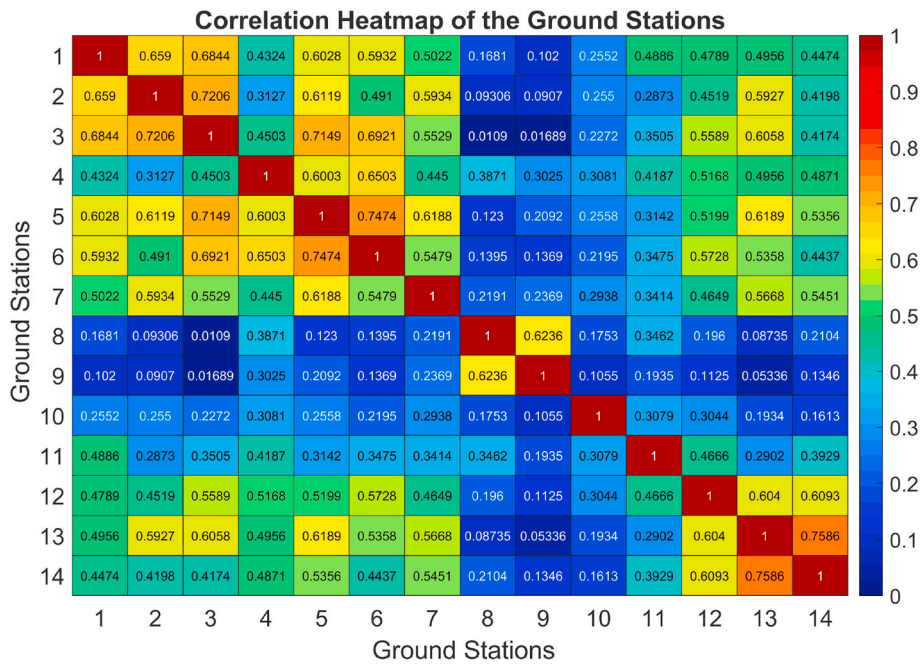


Fig. 3. Heatmap of the correlations of daily average NO2 concentration values among the fourteen stations over the period 2019–2020. The station numbers are as mentioned in Table 2.

model, and seasonal MA are presented in Eqs. (10)–(13), respectively [44].

$$\varphi_P(B^m)\varphi_P(B)(1 - B^m)^D(1 - B)^d y_t = \theta_Q(B^m) \theta_q(B)\omega_t \tag{8}$$

$$B^k y_t = y_{t-k} \tag{9}$$

$$\varphi(B) = 1 - \varphi_1(B) - \varphi_2 B^2 - \dots - \varphi_p B^p \tag{10}$$

$$\theta(B) = 1 + \theta_1(B) + \theta_2 B^2 + \dots + \theta_q B^q \tag{11}$$

$$\varphi_P(B^m) = 1 - \varphi_1(B^m) - \varphi_2 B^{2m} - \dots - \varphi_P B^{Pm} \tag{12}$$

$$\theta_Q(B) = 1 + \theta_1(B^m) + \theta_2 B^{2m} + \dots + \theta_Q B^{Qm} \tag{13}$$

where.

- y_t : nonstationary time series;
- ω_t : Gaussian white noise;
- $\varphi_P(B^m)\varphi_P$: seasonal autoregression polynomial;
- $\varphi_P(B)$: non-seasonal autoregression polynomial;
- $\theta_Q(B^m)$: seasonal moving average polynomial;
- $\theta_q(B)$: Non-seasonal moving average polynomial;
- D : seasonal difference term;
- B : Backshift operator.

3.2. Machine learning methods

3.2.1. Long short-term memory (LSTM)

Long short-term memory (LSTM) networks are recurrent neural networks (RNNs) used for time-series forecasting. They process input data over time and update the network state that holds the required information from the previous time steps to be used as an input for the prediction of the next step.

The architecture of a vanilla LSTM contains a set of recurrently associated sub-networks called memory blocks that maintain their state over time and regulate information flow through non-linear gating units. The LSTM unit consists of a cell, input gate, output gate,

forget gate, input signal $x^{(t)}$, output $y^{(t)}$, activation functions (sigmoid and tanh), and peephole connections. The cell remembers values over time intervals, and the three gates control the flow of information linked to the cell. The output of the block is recurrently connected to the next block input that passes through the cell and gates. The forward pass in the recurrent neural system is described by the process between the block input and output in the input gate, forget gate, updated cell value, output gate, and block output. The block input is responsible for updating the input component that combines the current input $x^{(t)}$ and output of the LSTM unit $y^{(t-1)}$ in the last step, as presented in Eq. (14).

$$z^{(t)} = g(W_z x^{(t)} + R_z y^{(t-1)} + b_z) \tag{14}$$

where W_z and R_z are the weights associated with $x^{(t)}$ and $y^{(t-1)}$, respectively, b_z represents for the bias weight vector, and g is the hyperbolic tangent (\tanh) activation function. The input gate is updated by combining the current input $x^{(t)}$, the output of the LSTM unit $y^{(t-1)}$, and the cell value $c^{(t-1)}$ in the last iteration, as illustrated in Eq. (15).

$$i^{(t)} = \sigma(W_i x^{(t)} + R_i y^{(t-1)} + P_i \odot c^{(t-1)} + b_i) \tag{15}$$

where W_i , R_i , and P_i are the weights associated with $x^{(t)}$, $y^{(t-1)}$, and $c^{(t-1)}$, respectively, while b_i represents the bias vector associated with this component, \odot denotes the point-wise multiplication of two vectors, and σ represents a logistic sigmoid activation function $\sigma(x) = \frac{1}{1+e^{-x}}$. The forget gate determines which information should be eliminated from the previous cell states $c^{(t-1)}$. Therefore, the activation values $f^{(t)}$ of the forget gates at time step t are computed based on the current input $x^{(t)}$, output $y^{(t-1)}$, and state $c^{(t-1)}$ of the memory cells at the previous time step ($t-1$) and the bias term b_f of the forget gate as follows:

$$f^{(t)} = \sigma(W_f x^{(t)} + R_f y^{(t-1)} + P_f \odot c^{(t-1)} + b_f) \tag{16}$$

where W_f , R_f , and P_f are the weights associated with $x^{(t)}$, $y^{(t-1)}$, and $c^{(t-1)}$, respectively, while b_f represents the bias weight vector. The current cell state value $c^{(t)}$ is updated after selecting the candidate values $z^{(t)}$ and input gate $i^{(t)}$ of the input gates and removing information from the previous cell states $c^{(t-1)}$ using the forget gate.

$$c^{(t)} = z^{(t)} \odot i^{(t)} + c^{(t-1)} \odot f^{(t)} \tag{17}$$

The output gate is computed by combining the current input $x^{(t)}$, output of the LSTM unit $y^{(t-1)}$ and cell value $c^{(t-1)}$ in the last iteration.

$$o^{(t)} = \sigma(W_o x^{(t)} + R_o y^{(t-1)} + P_o \odot c^{(t-1)} + b_o) \tag{18}$$

where W_o , R_o , and P_o are the weights of $x^{(t)}$, $y^{(t-1)}$, and $c^{(t-1)}$, respectively, and b_o represents the bias weight vector. The final step is the block output, which combines the current cell value $c^{(t)}$ with the current output gate value, as follows:

$$y^{(t)} = g(c^{(t)} \odot o^{(t)}) \tag{19}$$

The weights in the network were learned through backpropagation over time. The cell state $c^{(t)}$ receives gradients from $y^{(t)}$, and the gradients are gathered before being backpropagated to the current layer. The changes in the gradients in the last iteration represent the loss function [47–49].

3.2.2. Nonlinear autoregressive neural network (NAR-NN)

The NAR neural network is a dynamic neural network that predicts a time series. It has a feedback connection linking several layers of the network: the input layer, hidden layer, and output layer. NAR is expressed by the following equation, where $y(t)$ is the output value at time (t) and (d) is the delay order.

$$y(t) = f(y(t-1), y(t-2), y(t-3), \dots, y(t-d)) \tag{20}$$

The training function of the NAR neural network is the (LM) algorithm. NAR uses a hyperbolic tangent sigmoid transfer function for the hidden layer and a linear transfer function for the output layer. The NAR-NN test can be carried out using open and closed loops [47,50–53].

3.3. Evaluation criteria

All forecasted NO2 concentrations for the 14 ground stations that were obtained by applying the statistical and machine learning methods were evaluated based on the mean absolute percentage error (MAPE) evaluation criteria, which is one of the most commonly used metrics in forecasting [54]. The MAPE is determined by the following equation:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|A_t - F_t|}{A_t} \times 100 \tag{21}$$

where n is the number of observations or predictions, A_t is the actual value, and F_t is the forecast value.

For interpretation, a MAPE value lower than 10% is considered very good, <20% is good, <50% is acceptable, and >50% is not good [55].

4. Results and discussion

4.1. Experimental setup

The first 23 months of 2019–2020 (700 days) were used for training each of the four methods (LSTM, NAR-NN, ARIMA, and SARIMA). The trained models were then used to predict the NO₂ concentration values during the testing period of December 2020 (31 days). The forecasted values of NO₂ concentration during the 31 testing days were evaluated against the actual values using the MAPE (Eq. (21)), with $n = 31$.

The architecture and hyperparameters used in each method are listed in Table 3. These were empirically tuned to yield reasonably good results.

Notes.

- (i) For the LSTM, we also attempted a window size of 14 but did not obtain significantly different results.
- (ii) For SARIMA, we applied 14 seasonal variables but did not obtain significantly different results.
- (iii) More than 14 seasonal variables could not be applied because of limitations in computing resources.
- (iv) For LSTM, to avoid the problem of data leakage, the training data were only obtained for 693 days (instead of 700) because a 7-day sequence window is used.

4.2. Open loop vs closed loop forecasting

There are two basic modes of forecasting: open-loop and closed-loop. The open loop uses the actual input value from the previous steps to predict the value of the next step [56]. In other words, it represents forecasting with a one-unit time lag. When the time unit is a day, it is known as “one day-ahead” forecasting [57]. In the machine learning literature, this process is known as either “interleaved test-then-train” or “prequential” evaluation [58]. If testing data are obtained for T number of days, then we must repeat this interleaved test-then-train process T times, as shown in Fig. 4(a).

In our case, using the open-loop architecture, 700 days (Jan 1, 2019, to Nov 30, 2020) were used for training and one day (Dec 1, 2020) was used for testing. Then, 701 days (Jan 1, 2019, to Dec 1, 2020) were used for training and one day (Dec 2, 2020) was used for testing.

In contrast, closed-loop forecasting uses previous predictions as an input to forecast the next step [56]. Actual input values for the testing data were not available throughout the process. Testing data for T number of days is equivalent to carrying out multiple forecasts from “one day-ahead,” “two days-ahead,” etc., up to “ T days-ahead” one after another, as demonstrated in Fig. 4(b).

In our case, closed-loop forecasting was performed for 1 d to 31 d-ahead forecasts on the testing data. Usually, the closed-loop architecture is much faster than the open-loop architecture but less accurate. This phenomenon was confirmed in our experimental results, as shown in the section below.

4.3. Forecasting results and interpretations

4.3.1. MAPE results

The forecasting results in terms of the MAPE values of the four methods (LSTM, NAR-NN, ARIMA, and SARIMA) for both the open and closed loops are listed in Table 4, and the average MAPE values for the four methods, together with the interpretations according to Allwright [55] are listed in Table 5 [55].

Table 6 shows the p-values for the pairwise differences in MAPE values across the 14 stations using the two-tailed Wilcoxon signed-rank test [59]. For the open loop, only the differences between LSTM and ARIMA and LSTM and SARIMA were significant, while the rest were not. Therefore, a clearly optimal model was not determined for the open-loop setup. For the closed loop, four out of the six

Table 3
Architecture and hyperparameters used for the four methods.

Method	Architectures and Hyperparameters
LSTM	Sequence window size = 7 Unidirectional 1 LSTM layer with 64 neurons 1 hidden layer with 32 neurons Number of epochs = 50
NAR-NN	1 hidden layer with 4 neurons Number of epochs = 350
ARIMA	(p, d, q) = (1, 1, 1)
SARIMA	(p, d, q) = (1, 1, 1) (P, D, Q) _m = (1, 1, 1) ₇

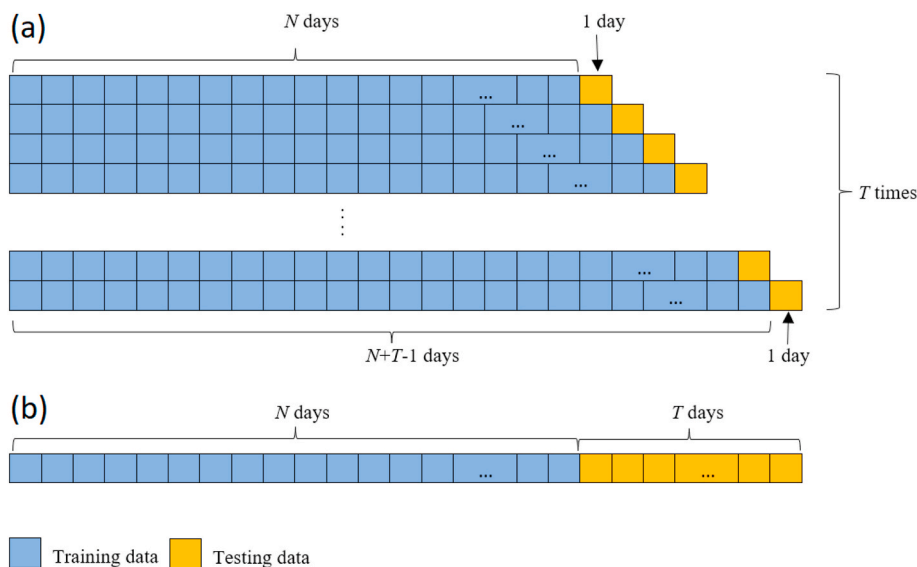


Fig. 4. Comparison of the two forecasting architectures with N days for training and T days for testing: (a) open loop and (b) closed loop.

Table 4

MAPE values for each model using the open and closed loop architectures to predict NO2 concentrations in December 2020.

Ground Station		Open loop				Closed loop			
		LSTM	NAR-NN	ARIMA	SARIMA	LSTM	NAR-NN	ARIMA	SARIMA
1	Hamdan Street	13.36	12.68	12.51	11.52	14.7	15.88	15.42	15.06
2	Khadejah School	37.53	31.69	37.37	33.88	39.12	40.00	46.54	44.13
3	Khalifa School	34.68	33.82	32.11	33.04	34.16	37.75	46.15	45.67
4	Baniyas School	15.87	18.20	15.13	13.82	15.21	17.88	17.86	16.60
5	Bain Al Jessrain	25.38	24.26	25.41	24.74	25.50	31.76	33.09	33.31
6	Al Mafraq	12.06	11.99	11.03	8.99	11.96	15.39	17.35	16.12
7	Bida Zayed	33.42	32.50	33.88	34.08	33.43	33.08	39.80	40.00
8	Gayathi School	22.70	22.79	22.28	23.15	22.54	19.57	25.35	25.70
9	Ruwais	28.15	27.91	28.31	29.50	26.65	28.29	31.74	32.34
10	Liwa	9.05	8.00	8.71	8.81	8.90	9.99	11.95	12.08
11	Al Quaa	23.64	26.57	23.28	22.97	24.50	32.30	36.90	35.63
12	Sweihaan	21.81	23.86	21.49	19.27	24.05	23.81	23.80	23.85
13	Al Tawia	30.54	29.44	29.39	29.63	30.72	24.71	24.62	24.39
14	Zakher	26.19	27.29	24.18	23.77	25.68	23.82	24.61	24.77
Average		23.88	23.64	23.22	22.66	24.08	25.30	28.23	27.83

pairs (except for LSTM vs. NAR-NN and ARIMA vs. SARIMA) were significantly different, and this finding implies that the machine learning methods (LSTM/NAR-NN) were better than the statistical methods (ARIMA/SARIMA) in the closed-loop setup.

For the overall comparison, we found that the average MAPE values of the open loop were significantly lower than those of the closed loop. This was observed at 12 out of the 14 stations (with the exception of the two stations in the Al Ain Region). To confirm this phenomenon, we conducted two statistical tests across 14 stations. The one-tailed paired *t*-test [60] resulted in a *p*-value of 0.00266 while the one-tailed Wilcoxon signed-rank test [59] resulted in a *p*-value of 0.00453, both of which are significantly smaller than 0.05.

In addition to the analysis of the results with the MAPE evaluation criteria, we have also carried out an additional analysis using the Root-Mean-Square Error (RMSE) criteria [61]. The results are presented in Section S.1 of the supplementary materials.

4.3.2. Lowest, medium, and highest cases

For each loop type, among the 14 ground stations, three were selected to represent the lowest, medium, and highest MAPE values.

The lowest MAPE values for both the open and closed loops were obtained at the Liwa station, which is located in the rural area of the Al Dhafra Region and had MAPE values of 8.64 and 10.73 respectively. This means that the concentration of NO2 in Liwa was well predicted and Liwa was unaffected by the fluctuation of NO2 during the COVID-19 lockdown. The emissions of NO2 in Liwa were approximately the same before and after the COVID-19 lockdown.

The second station was selected to present the medium MAPE values among all the other ground stations in both open and closed loops architectures, which were ranked as the seventh best MAPE values in each loop type. The Al Quaa station (rural area in Al Ain Region) was selected for the open loop, and the Zakher station (urban area in Al Ain Region) was selected for the closed loop, and their

Table 5

Average MAPE values (LSTM, NAR-NN, SRIMA, and SARIMA) for each ground station using the open and closed loop architectures of predicting NO₂ concentrations in December 2020. The interpretations are based on Allwright [55].

Ground Station	Open Loop		Closed Loop		Difference in avg. MAPE (Closed loop – Open loop)
	Average MAPE	Interpretation	Average MAPE	Interpretation	
1 Hamdan Street	12.52	Good	15.27	Good	2.75
2 Khadejah School	35.12	Acceptable	42.45	Acceptable	7.33
3 Khalifa School	33.41	Acceptable	40.93	Acceptable	7.52
4 Baniyas School	15.76	Good	16.89	Good	1.13
5 Bain Al Jessrain	24.95	Acceptable	30.92	Acceptable	5.97
6 Al Mafraq	11.02	Good	15.21	Good	4.19
7 Bida Zayed	33.47	Acceptable	36.58	Acceptable	3.11
8 Gayathi School	22.73	Acceptable	23.29	Acceptable	0.56
9 Ruwais	28.47	Acceptable	29.76	Acceptable	1.29
10 Liwa	8.64	Very good	10.73	Good	2.09
11 Al Quaa	24.12	Acceptable	32.33	Acceptable	8.21
12 Sweihan	21.61	Acceptable	23.88	Acceptable	2.27
13 Al Tawia	29.75	Acceptable	26.11	Acceptable	−3.64
14 Zakher	25.36	Acceptable	24.72	Acceptable	−0.64
Average	23.35	Acceptable	26.36	Acceptable	3.01

Table 6

P-values for the pair-wise differences in MAPE values across the 14 stations using the two-tailed Wilcoxon signed-rank test: (a) open loop architecture and (b) closed loop architecture. Statistically significant p-values (<0.05) are highlighted in bold.

	(a) Open Loop Architecture				(b) Closed Loop Architecture		
	NAR-NN	ARIMA	SARIMA		NAR-NN	ARIMA	SARIMA
LSTM	0.70394	0.01016	0.01314	LSTM	0.17702	0.01314	0.01552
NAR-NN		0.33204	0.27134	NAR-NN		0.00758	0.01108
ARIMA			0.17702	ARIMA			0.19706

MAPE values were 24.12 and 24.72, respectively.

For both loop types, the highest MAPE values were measured at the Khadejah School station, which is located in an urban area in the Abu Dhabi Region, and it had a MAPE of 35.12 with the open-loop architecture and 42.44 with the closed-loop architecture.

In general, the MAPE values using the open-loop architecture are better than those of the closed loop architecture, as presented in Table 5, Figs. 5 and 6, which is also confirmed by the two statistical tests discussed in the last paragraph of the above section. We can observe that for the closed loop, NAR-NN and SARIMA produced forecasts for all 31 days with only small variations while ARIMA generated forecasts with virtually no variations after a few days. These phenomena can be observed at all three representative stations in Fig. 6.

4.3.3. Correlation analysis

We also investigate why the forecasting accuracies (in terms of the MAPE) varied among the 14 stations. First, we arranged the average NO₂ concentration values for December 2020 in ascending order and then tabulated and plotted the average MAPE values of the open- and closed-loop forecasts, Table S2(a), in the supplementary materials, and its corresponding Fig. 7(a). We noticed that the mean NO₂ concentration values and MAPE were poorly correlated when measured using the Pearson correlation coefficient [43]. The correlation coefficients are −0.36413 and −0.25975, respectively for the open- and the closed-loop setups.

The relative standard deviation (RSD), which is also known as the coefficient of variation [62], is the standard deviation (SD) expressed as a percentage of the mean. It is calculated as follows (Eq. (22)):

$$RSD = \left(\frac{SD}{mean} \right) \times 100 \quad (22)$$

We now arrange the RSD values of the NO₂ concentration for December 2020 in ascending order. The average MAPE values of the open- and closed-loop forecasts were tabulated and plotted, as shown in Table S2(b), in the supplementary materials, and its corresponding Fig. 7(b). The RSD values were highly correlated with the MAPE values for both the open- and closed-loop setups (0.92782 and 0.94443, respectively). This reflects a very general phenomenon in forecasting: more accurate forecasts are made when the data exhibit fewer variations while less accurate forecasts are made when the data exhibit higher variations.

5. Conclusions

This study compared the performance of the statistical and machine learning models ARIMA, SARIMA, NAR, and LSTM using the MAPE evaluation criteria. In addition, the models were compared in both open- and closed-loop architectures. The results show that

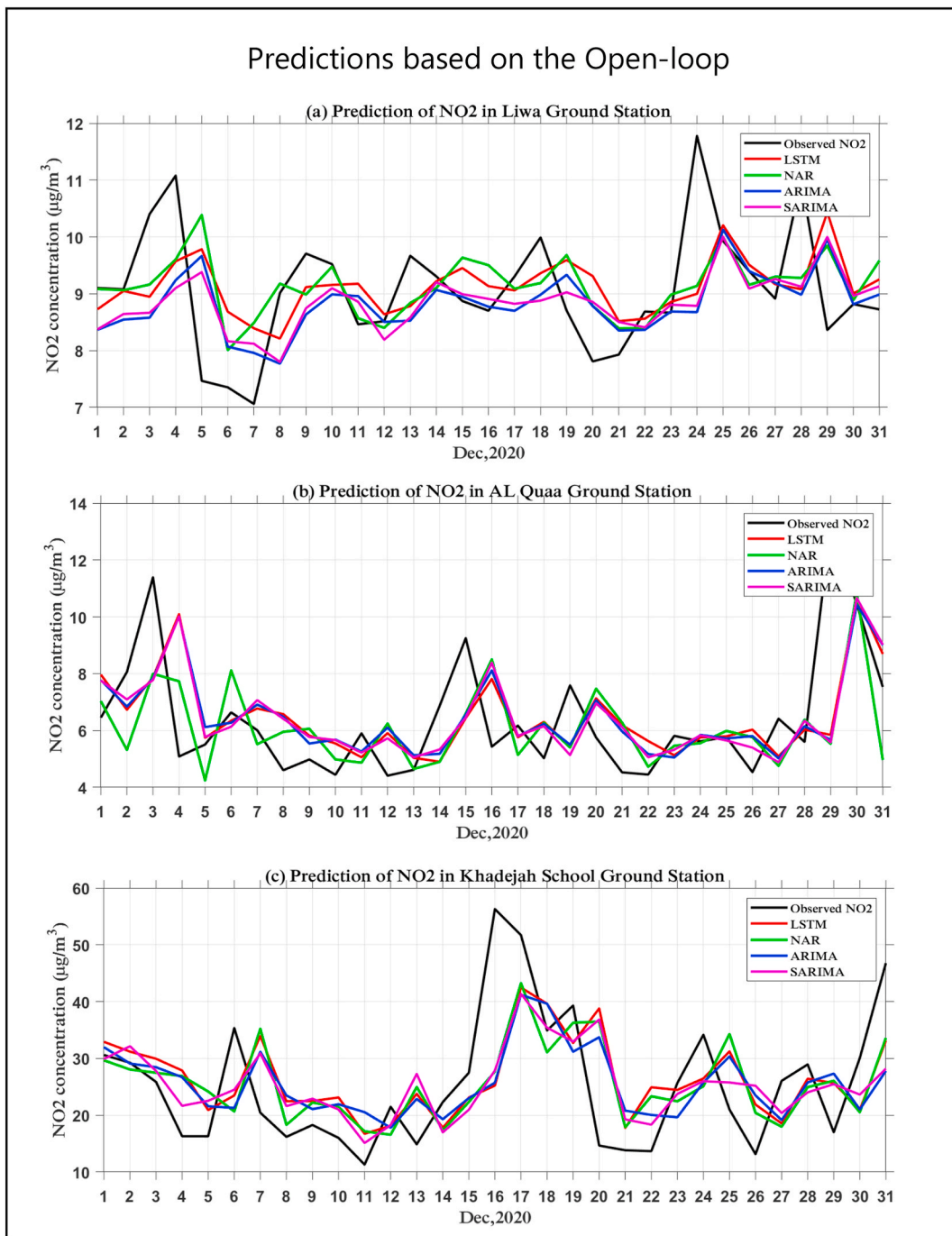


Fig. 5. Prediction of NO₂ concentrations during Dec 2020 using the open loop architecture for the ARIMA, SARIMA, NAR, and LSTM models. Predicted NO₂ concentrations in (a) Liwa, (b) Al Quaa, and (c) Khadejah School.

the open loop, which takes the actual value as an input to the next forecasting step, yields statistically better MAPE values than the closed loop, which feeds the previous output of the prediction value as an input to the next step. However, the MAPE values at two stations (Al Tawia and Zakher) in urban and suburban areas of Al Ain Region were lower (better) when using a closed loop compared to the open loop. This might be because of the nature of the desert region or effects of different activities. An example of fluctuations in the activities was found in Al Quaa and Zakher. The MAPE value in Al Quaa using the open loop was better than that of the closed loop, while this finding was reversed in Zakher. The prediction of NO₂ in Liwa was the best when using either the open or closed loop. This could be because Liwa is a rural area unaffected by human activities and thus unaffected by the COVID-19 lockdown. The prediction

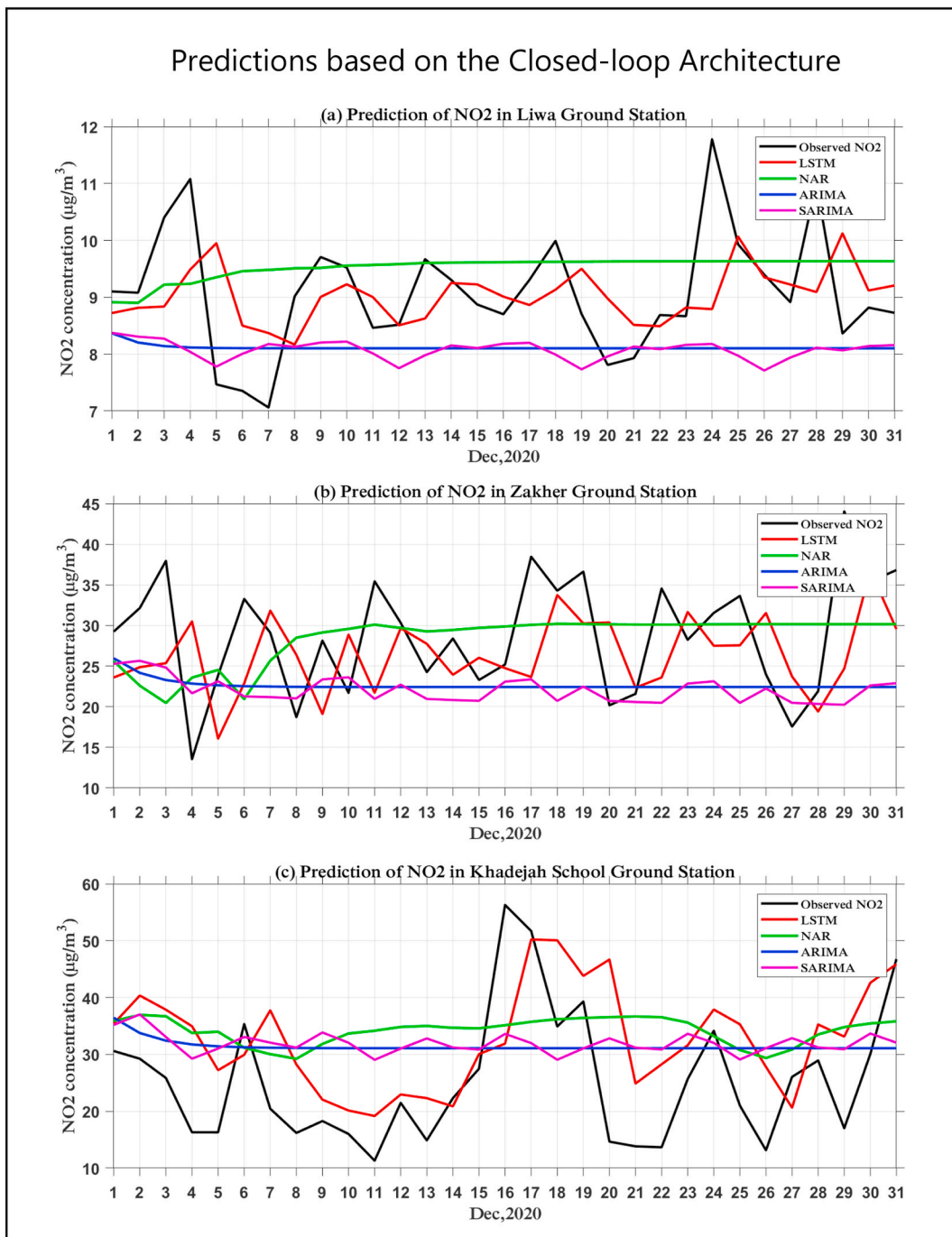


Fig. 6. Prediction of NO₂ concentrations during Dec 2020 using the closed loop architecture for the ARIMA, SARIMA, NAR, and LSTM models. Predicted NO₂ concentration in (a) Liwa, (b) Zakher, and (c) Khadejah School.

for Khadejah School (in the urban area of Abu Dhabi with many NO₂ fluctuations) was the worst using both open and closed loops. A comparison of the four methods (LSTM, NAR-NN, ARIMA, and SARIMA) did not identify a clearly optimal method for the open-loop setup. However, the machine learning methods (LSTM/NAR-NN) were generally better than the statistical methods (ARIMA/SARIMA) for the closed-loop setup. Finally, we showed that the MAPE is highly correlated with the RSD for both loop types.

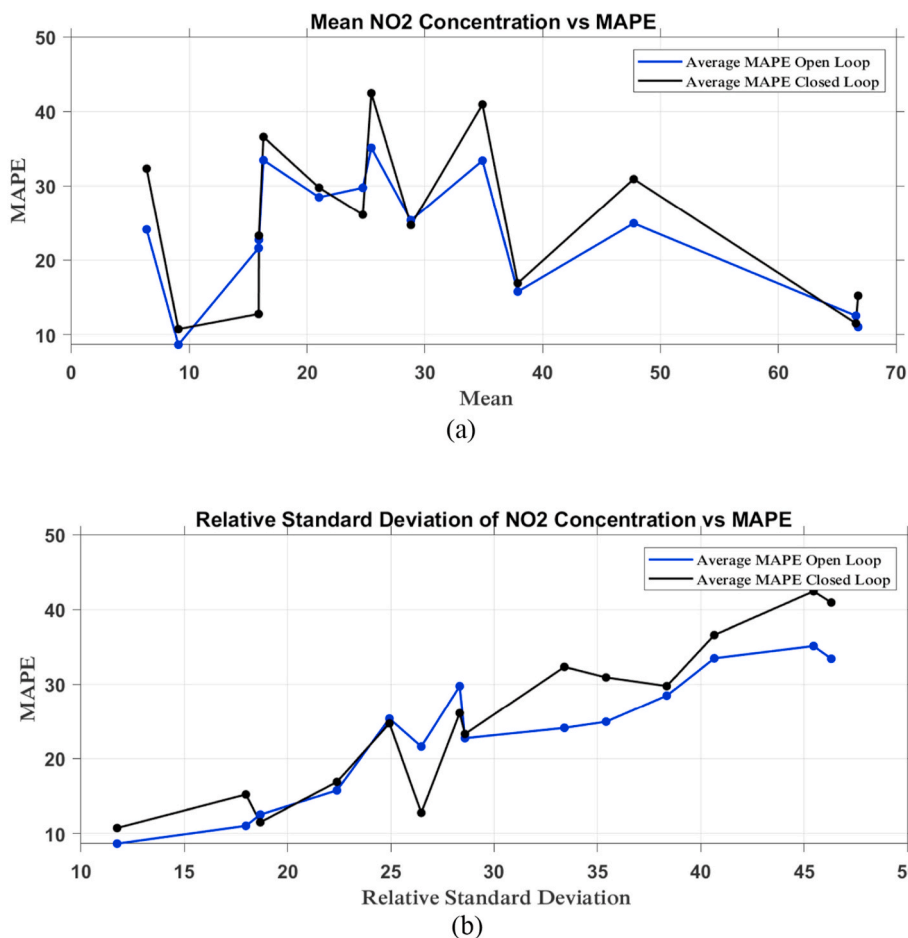


Fig. 7. Graphical presentations of the (a) mean NO₂ concentration vs MAPE (open and closed loops) and (b) RSD of the NO₂ concentration vs MAPE (open and closed loops) across the fourteen stations in December 2020.

Data availability

The ground station data used in the study to support findings were made available by Environment Agency – Abu Dhabi specifically for the study under a license agreement. Hence, the data are not publicly available and must be requested directly from the Environment Agency – Abu Dhabi by writing to customerhappiness@ead.gov.ae.

Acknowledgements

This work was supported in part by the Environment Agency – Abu Dhabi.

Appendix B. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.heliyon.2022.e12584>.

References

- [1] X. Yang, Y. Wang, C. Zhao, H. Fan, Y. Yang, Y. Chi, L. Shen, X. Yan, Health risk and disease burden attributable to long-term global fine-mode particles, *Chemosphere* 287 (2022), 132435.
- [2] J.L. Santiago, E. Rivas, A.R. Gamarra, M.G. Vivanco, R. Buccolieri, A. Martilli, Y. Lechón, F. Martín, Estimates of population exposure to atmospheric pollution and health-related externalities in a real city: the impact of spatial resolution on the accuracy of results, *Sci. Total Environ.* 819 (2022), 152062.
- [3] H. Fan, C. Zhao, Y. Yang, A comprehensive analysis of the spatio-temporal variation of urban air pollution in China during 2014–2018, *Atmos. Environ.* 220 (2020), 117066.

- [47] M.H. Beale, M.T. Hagan, H.B. Demuth, *Neural Network Toolbox User's Guide*, in: *The MathWorks*, 2010, pp. 77–81.
- [48] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.* 9 (8) (1997) 1735–1780.
- [49] G. Van Houdt, C. Mosquera, G. Nápoles, A review on the long short-term memory model, *Artif. Intell. Rev.* 53 (8) (2020) 5929–5955.
- [50] S.S. Chaharborj, M.M. Borazjani, Y. Gheisari, Predicting climatic meteorological parameters by using the artificial dynamics neural networks: case study, Bushehr City, *J. Comput. Sci. Comput. Math.* 6 (2016) 113–119.
- [51] S. Ferlito, M. Atrigna, G. Graditi, S. De Vito, M. Salvato, A. Buonanno, G. Di Francia, Predictive Models for Building's Energy Consumption: an Artificial Neural Network (ANN) Approach, in: *Proceedings of the 2015 XVIII AISEM Annual Conference*, 2015, pp. 1–5.
- [52] B. Liu, Y. Zhang, Calibration of miniature air quality detector monitoring data with PCA–RVM–NAR combination model, *Sci. Rep.* 12 (2022) 9333.
- [53] D. Perez Bello, M.P. Natali, A. Meza, Comparison of adaptive neuro-fuzzy inference system and recurrent neural network in vertical total electron content forecasting, *Neural Comput. Appl.* 31 (12) (2019) 8411–8422.
- [54] S. Makridakis, M. Hibon, Accuracy of forecasting: an empirical investigation, *J. Roy. Stat. Soc.* 142 (1979) 97–125.
- [55] S. Allwright, What is a good MAPE score?, Retrieved from, <https://stephenallwright.com/good-mape-score/>, 2021.
- [56] MathWorks, Time series forecasting using deep learning, Retrieved from, <https://www.mathworks.com/help/deeplearning/ug/time-series-forecasting-using-deep-learning.html>, 2022.
- [57] G. Draluz, D. Mazur, M. Gołębowski, L. Gołębowski, One Day-Ahead Forecasting at Different Time Periods of Energy Production in Photovoltaic Systems Using Neural Networks, in: *Proceedings of the 2018 International Symposium on Electrical Machines (SME)*, 2018, pp. 1–5.
- [58] J. Gama, R. Sebastiao, P.P. Rodrigues, On evaluating stream learning algorithms, *Mach. Learn.* 90 (2013) 317–346.
- [59] Janez Demsar, Statistical comparisons of classifiers over multiple data sets, *J. Mach. Learn. Res.* 7 (1) (2006) 1–30.
- [60] Bernice D. Mowery, The paired t-test, *Pediatr. Nurs.* 37 (6) (2011) 320.
- [61] W. Wang, Y. Lu, Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model, in: *IOP Conference Series: Materials Science and Engineering*, vol. 324, 2018, 012049 no. 1.
- [62] C.E. Brown, Coefficient of Variation, in: *Applied Multivariate Statistics in Geohydrology and Related Sciences*, 1998, pp. 155–157.