Palestine Journal of Mathematics

Vol 13(Special Issue I)(2024) , 209–221

AIR QUALITY PREDICTING USING LSTM RECURRENT NEURAL NETWORK IN TANGIER: A COMPARATIVE ANALYSIS DETECTING OZONE CONCENTRATION PEAKS

Nisrine Marrakchi, Amal Bergam, Hanane Fakhouri and Kenza Khomsi

MSC 2010 Classifications: Primary 20M99, 13F10; Secondary 13A15, 13M05.

Keywords and phrases: : Air quality forecasting, Artificial Neural Network, recurrent neural networks (RNN), Long-Short-Term Memory (LSTM), Ozone (O3).

Abstract Air quality forecasting has highly become critical given the serious negative effects on human health, caused by exposure to atmospheric pollutants in urban air. One of the pollutants associated with air pollution issues in Tangier city is troposphere ozone (O3), which is becoming more and more concerned. Predicting daily concentrations of this pollutant can be greatly beneficial in reducing possible risks. This study investigates the potential of using two types of recurrent neural networks (RNN): simple RNN and Long-Short-Term Memory (LSTM) model to forecast ozone concentrations in the Tangier area. In each experiment utilizing these two algorithms, we adjusted the input data to examine the relation between predictors and performance. When evaluating our forecasting model, the predicted and observed values were displayed graphically. The error was also measured using a number of evaluation metrics, including Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE) and the Index of Agreement. The computational results demonstrate that both approaches functioned as intended, although the LSTM model outperforms a simpleRNN architecture in forecasting ozone (O3) concentrations.

1 Introduction

Nowadays, forecasting air pollution has become a significant concern. Air quality directly affects both our health and the ecosystem, especially in large cities. The numerous changes seen in the biosphere, oceans, and atmosphere are mainly a consequence of human contribution. Therefore, there is no doubt that the increase in the majority of pollutants in the ambient air around the industrial region [1]. Tangier, north of Morroco, has seen significant growth in a number of industrial sectors. However, the environment and public health have been negatively impacted by this development. Most of Tangier's air pollution issues are caused by a number of pollutants, including tropospheric ozone (O3) [2].

Ozone in the troposphere is produced by a unique mechanism since it is not generated directly by people's activities: it is a secondary pollutant that results from photochemical reactions within the atmosphere that need sun energy [3]. Furthermore, multiple lines of scientific evidence support the claim that there are lots of elements available that contribute to the generation of O3 molecules, including nitrogen dioxide (NO2), volatile organic compounds (VOC) or the carbon monoxide (CO) and in the presence of Ultra-Violet radiation [1]. There is also an elevated level of confidence that human changes and an increase in precursor emissions over the last decade have led to a rise in tropospheric ozone [4].

On plants, ozone has an adverse influence. As a strong oxidant, it destroys plant cells by interacting with their surface elements, which is typically observed on the leaves where necrosis can occur. Ozone may also enter through the small airways in humans, which can lead to asthma episodes and coughing. In addition, it has been shown that exposure to high concentrations of ozone increases death and hospitalization [5].

At the national level, more assistance is required for the successful implementation of air pollution control and the reduction of precursor emissions that cause ozone. In Morroco, legislation

governing air quality is being developed, but we still don't have a solid forecasting system to deal with the escalating effects of air pollution and climate change. As a result, the creation of prediction models and the fusion of innovative technology with the particular conditions in Tangier region can close the air quality monitoring gaps and allow authorities to put ozone precursor mitigation measures into action.

Literature review

The precision of the prediction is an essential part of air quality forecasting. As a result, great efforts have been made to improve the accuracy of this measurement method. In this work's paper, the accuracy is achieved by applying some error measuring techniques, which include RMSE, MAPE and the index of agreement, and then comparing the actual pollution levels with the expected results produced by our proposed models. The following section reviews some of the associated research studies. The deterministic and statistical models are now two distinctive types of models as a result of recent advances in air quality approaches [6]. On the one hand, statistical models need to identify the underlying links between the various components that affect air quality to be able to provide accurate predictions. On the other hand, deterministic models work on the basis of solving differential equations that describe the atmospheric process [3].

There is no one perfect solution that can address all time series forecasting concerns. However, each problem could be resolved in a different way [8]. The Moving Average (MA) is considered one of the most simple forecasting techniques for time series without an identifiable seasonal trend [9].A more sophisticated version of MA known as Autoregressive Integrated Moving Average (ARIMA) has been employed in multiple articles, and the results show that the latter is credible for the prediction of time series [10].

Some studies have compared several alternative regression methods to predict daily ozone, these estimations are frequently based on statistical correlations between meteorological variables and levels of air pollution in the environment [12]. Furthermore, numerous researchers have employed multivariate linear regression models for this objective, and they can produce reliable findings. However, the interactions between pollution and meteorological factors tend to be complicated and nonlinear, particularly for ozone characteristics, which neural networks techniques may be better suited to model [13].

A different approach to prediction is the use of artificial neural networks (ANN). The end result of a research investigation that contrasts artificial neural networks with conventional approaches, such as exponential smoothing, the Box-Jenkins ARIMA model, and multivariate regression, revealed that the neural network model is generally more precise than alternative methods and is capable of understanding dynamic nonlinear trends and seasonal patterns [14]. Additionally, Artificial Neural Networks (ANNs) have produced favorable forecasts of ozone levels one hour in advance using data locally obtained in Corsica [7].

Recurrent Neural Network (RNN) is another technique that has received a lot of attention recently. RNNs are essentially networks composed of loops, which allows them to remember past events. Therefore, they can be quite helpful in time series prediction [15]. Long-Short-Term Memory (LSTM) technology is among the distinctive varieties of RNN, but what makes it so unique is how the hidden layer is monitored. As an illustration, [16] offers a time series anomaly detection model that uses LSTM.

Research contributions

- This study combines information on ozone concentration with meteorological data. First, to better comprehend the time series, a different exploratory data analysis is also needed.
- To obtain an accurate forecast of daily ozone in the city of Tangier, we use certain innovative artificial neural network-based forecasting techniques, including Simple RNN and LSTM [17].
- Using index of agreement and other accuracy measuring techniques, the results are reviewed.
- Finally,the best prediction techniques are selected after a careful analysis of the outcomes.

2 Materials and Methods

2.1 Study area and data samples

Tangier city in northern Morocco is the second economic heart of the country and capital of the Tangier-Tetouan-Al Hoceima region. It is located between 35 ° 46′ 01″ north and 5 ° 48′ 00″ east, with an area of about 124 km^2, see Figure 1.

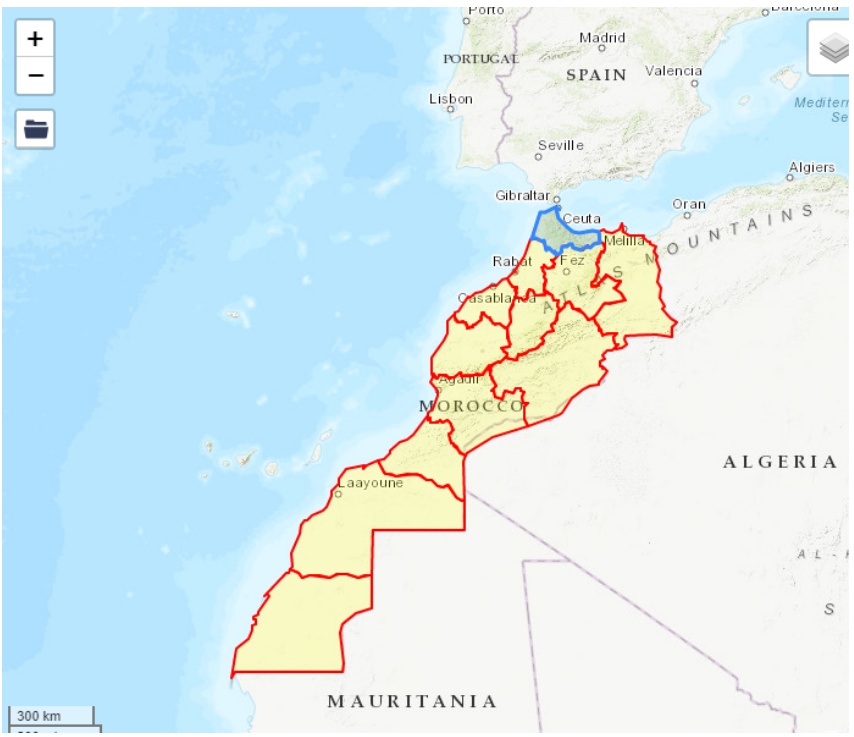


Figure 1: The location of Tangier, Morocco

There has been incredible business sector development in the Tangier region. However, this had an adverse effect on the ecology of the area, particularly on air quality and conditions for hygiene and population health. Gases and particulates emitted into the atmosphere by a variety of sources cause air pollution in Tangier. Road transportation and industrial sectors are the main sources of pollution in the atmosphere of Tangier [9]. In the current study, we used both the concentrations of air pollutant O3 and the meteorological factors. The air pollution data was measured and collected by the National Direction of Meteorology. The sampling period was from 01 January 2010 to 04 April 2014 in Tangier. We adopted K-Nearest Neighbors (KNN) imputation for interpolating the values that are lacking in the data and it offers higher precision and the finest performance.

2.2 Forecasting with Artificial Neural Networks

Several linked nodes, known as neurons, join together to create a neural network. Typically, a feedforward has three layers that make up its structure: an input layer, a hidden layer, and an output layer [13]. The number of attributes you wish to enter into the neural network is represented by the input layer's nodes. Additionally, the output nodes display the number of objects you desire to forecast or classify. Without forgetting that The hidden layer's function at this point will be to deal with the nonlinearity issue in particular. When calculating, every input x_i is multiplied by a specific weight w_i, and the sum of all weights is increased by a bias b specific to each neuron. in accordance with the formula 2.1:

$$\sum_{t=1}^{n} (w_i * x_i) + b \tag{2.1}$$

The activation function utilizes the determined quantity to obtain the y_i result based on Equation 2.2. The parameters of the neuron network are its weights and biases, and the algorithm used for training must determine them during a period of supervised learning. See Figure 2.

$$y_i = \sigma\left(\sum_{i=1}^{n}(w_i * x_i) + b\right) \tag{2.2}$$

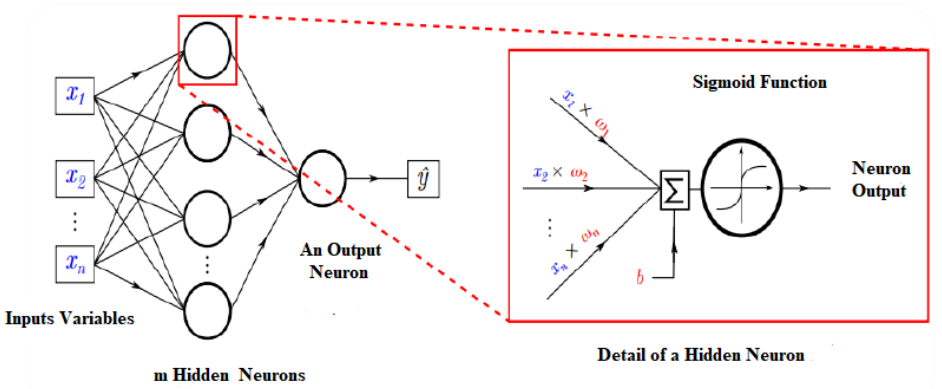


Figure 2: Schematic view of feed forward neural network

One of the frequent activation functions is the sigmoid that produces an output in the range between 0 and 1, according to this definition:

$$\sigma(x) = \frac{1}{1 + \exp(-x)}$$

The primary advantage of this function is that it is simple to derive. Its derivative, a crucial element of learning algorithms, is given by the following:

$$\frac{\partial \sigma}{\partial x}(x) = \sigma(x)(1 - \sigma(x))$$

After that, we trained our networks using the Gradient Descent Algorithm (DG), which allows us to determine the error gradient for each neuron in the network beginning with the last layer and ending up with a hidden layer. Gradient backpropagation adjusts the weights of the interconnections to decrease the MSE of the squared error between the output of our model and the target data [18].

$$\min \frac{1}{2}\sum_{i=1}^{N}\left(y_i - \sigma\left(\sum_{i=1}^{n} w_i * x_i + b\right)\right)^2$$

where, N stands for the dimension of the output vectors.

2.3 Forecasting with Recurrent Model Network

Time-series forecasting can benefit greatly from RNN's capacity to save previous events in its memory. The process goes like this: Data are sent from the input layer to the hidden layer, which consists of a repeating loop at the back. Then, the output is an expression of the preceding input, increased by the activation value of what was previously in hidden layers, as defined in Equation 2.3. RNN is shown in Figure 3 [20].

Each recurrent neural network receives a series of events as input $X = (x_1, x_2, .., x_N)$. The series of hidden states is defined by $h = (h_1, h_2, .., h_N)$ to produce the sequence of output vectors $y = (y_1, y_2, ..., y_N)$ from 1 to N iterations of t.

$$\begin{aligned} h_t &= H(w_{xh}x_t + w_{hh}h_{t-1} + b_h), \\ y_t &= w_{hy}h_t + b_y \end{aligned} \tag{2.3}$$

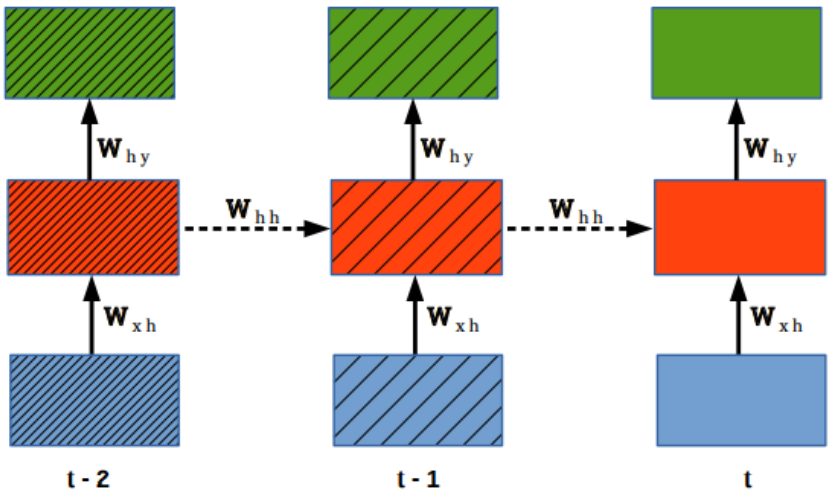


Figure 3: The process of a recurrent neural network (RNN)

where,

- N represents the overall number of input vectors,
- $w_{\alpha\beta}$ stands for weight matrix between the layers α and β,
- H is the hyperbolic tangent (tanh).

However, these networks have restrictions since, with accumulated computations, the error acquired with backpropagation of the gradient decreases towards zero or, less commonly, increases to infinity. The terms "vanishing gradient" and "gradient explosion" refer to these two issues, respectively.

2.4 Forecasting with LSTM

The Long-Short-Term Memory Network (LSTM) is a more powerful sequential network, often known as an RNN, that allows information persistence. It helps to resolve the issues of overcoming and vanishing RNN gradient [17].

The distinctive characteristic of LSTM is the way the hidden state is regulated, as shown in Figure 4. A "memory cell", which substitutes the hidden layer in LSTMs, is used to manage recurrence, which in the case of basic RNNs is ensured by an easy function known as the hyperbolic tangent (Tanh). The LSTM cell possesses a central node that stores the inside state (memories) of the cell in addition to certain "gates" [19].

$$
\begin{aligned}
i_t &= \sigma(w_{xi}x_t + w_{hi}h_{t-1} + w_{ci}h_{t-1} + b_i),\\
f_t &= \sigma\left(w_{xf}x_t + w_{h_f}h_{t-1} + w_{cf}c_{t-1} + b_f\right)\\
c_t &= f_t c_{t-1} + i_t \tanh\left(w_{xc}x_t + w_{h_c}h_{t-1} + b_c\right)\\
o_t &= \sigma\left(w_{xo}x_t + w_{h_o}h_{t-1} + w_{co}c_t + b_o\right)\\
h_t &= o_t \tanh\left(c_t\right)
\end{aligned}
$$

where,

- i_t represents Input gate
- f_t represents Forgot gate
- o_t represents Output gate
- h_t represents hidden state vector
- c_t represents cell state vector

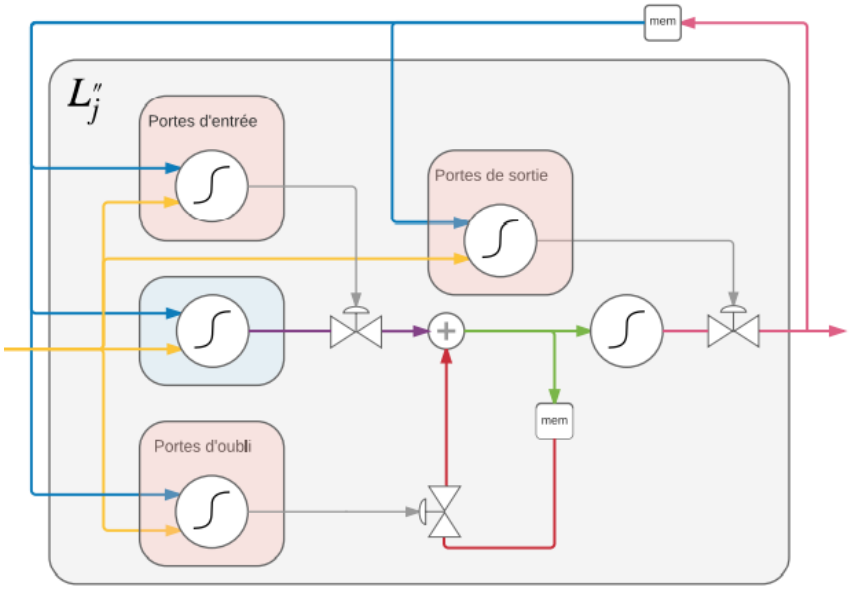


Figure 4: The process of Long Short-Term Memory Network (LSTM)

- w represents weight matrices
- b represents bias vector

It is desirable to resize the dataset to an interval from 0 to 1 in the initial stages of prediction due to the sensitive nature of LSTM. It is then possible to create the model. Additionally, it is beneficial to forecast future O3 concentrations after fitting. For this purpose, using the Keras package to generate an LSTM model requires converting the time series into a supervised learning problem by using the previous data, such as $t - 20$ as input and what was actually observed at time t as output [21].

2.5 forecasting error metrics

The assessments of forecasting time series reflect how effective the model is. There are many distinct approaches to evaluate models. Three widely utilized metrics are used in this stage, such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The expression of these assessment metrics shows that as the three values decrease, the model's prediction error also decreases.

Willmott introduced the index of agreement (d) as a standard technique to analyze the level of model prediction error [22]. This shows how the projected error and mean square error are related. A score of 0 denotes total disagreement, while a score of 1 denotes ideal agreement.

These metrics are calculated using the formulas (2.4-2.7).

$$\boldsymbol{MSE} = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2 \tag{2.4}$$

$$\boldsymbol{MAE} = \frac{1}{n}\sum_{t=1}^{n}(|y_t - \hat{y}_t|) \tag{2.5}$$

$$\boldsymbol{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{2.6}$$

$$\boldsymbol{d} = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}\left(|y_t - \bar{\hat{y}}_t| + |\hat{y}_t - \bar{\hat{y}}_t|\right)} \tag{2.7}$$

where,

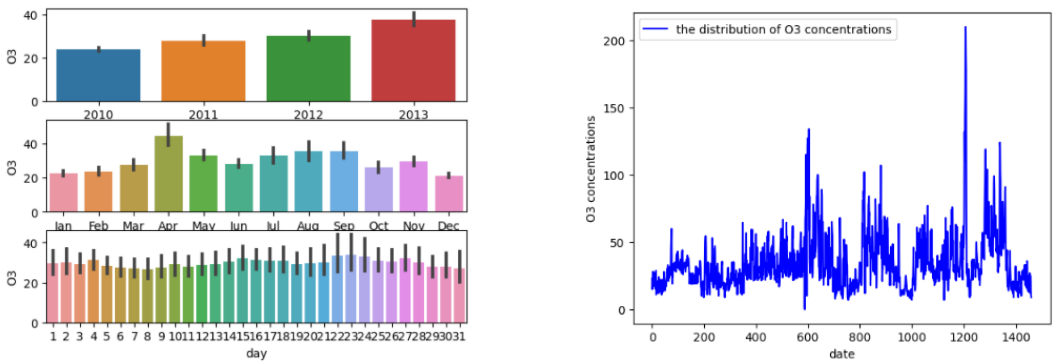
- y_t represents the observed value.
- $\hat{y}_t$ represents the forecasted value.
- $\bar{\hat{y}}_t$ represents the average of forecasted value.
- n represents the number of total observations.

3 Experiments

3.1 Prediction of time series

Forecasting time series is a significant subject in machine learning. A normal machine learning dataset is a collection of observations. A typical machine learning dataset is made up of several observations. However, a time series dataset differs because it includes a time dimension as additional information. In general, analyzing the ozone dataset and making predictions about the same pollutant are the two aims associated with this work.

The collection of data for the present investigation includes the number of daily O3 concentrations in the air from January 2010 to April 2014. The values are a count of concentrations, and there are 1556 observations. We split up the dataset into two groups: one for model development (dataset) and the other for validation (validation). The dataset was made up of observations from January 2010 to December 2013. Validation: data made between January and April 2014.



(a) (b)

Figure 5: Histogram Plot and Line plot of O3 distribution

In time series analysis and forecasting, visualization is crucial. Graphical representations of raw sample data can help diagnose the temporal structures that may affect the selection of the model. As well, we can observe that our series tends to increase over time, which justifies the choice of this pollutant, see Figure 5.

The analysis of time series is the main concern in order to achieve the objective of the study. Time series analysis is creating models that best capture or explain our actual data when attempting to comprehend what is causing them. This frequently means establishing hypotheses regarding the structure of the data that is divided into its four constituent parts, Level, Trend, Seasonality and Noise, as demonstrated in Figure 6.

These elements can aid us in formulating hypotheses about how they interact and how the observed time series will behave. They might also be a useful technique to forecast future values and provide us with the option to select the best prediction methodology. We are able to see that the O3 series behaves in a way that tends to rise over time, with a negligibly pronounced seasonality and a minor repeating pattern over time,and evidently some variation in our time series that the model is unable to account for, that is referred to as noise.

The graph of autocorrelation expresses the character and degree of the link between the observations and their lags. After the resulting graph, the autocorrelation value of the daily O3 dataset tends to zero, as shown in Figure 7. On observing the plot, it is clear that the autocorrelation value of the daily O3 dataset tends to zero, as seen in Figure 7. It appears that there is only a minimal relationship between the original series and its lag. This illustrates a minor correlation

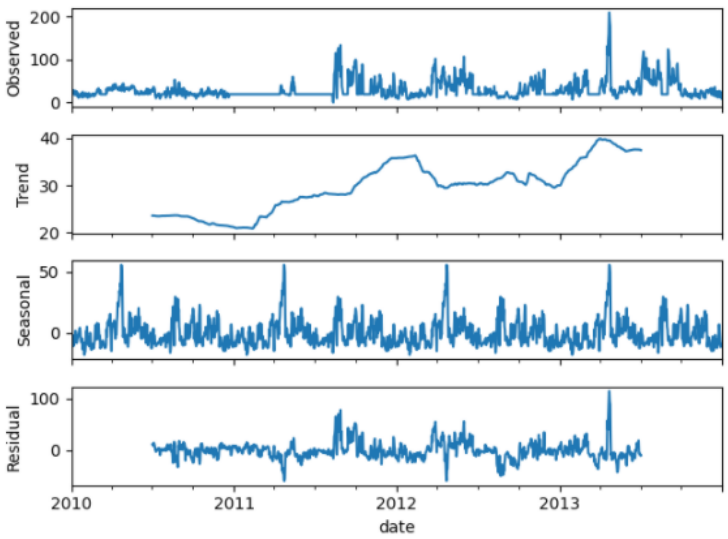


Figure 6: Decomposing of the O3 time series into four constitution components

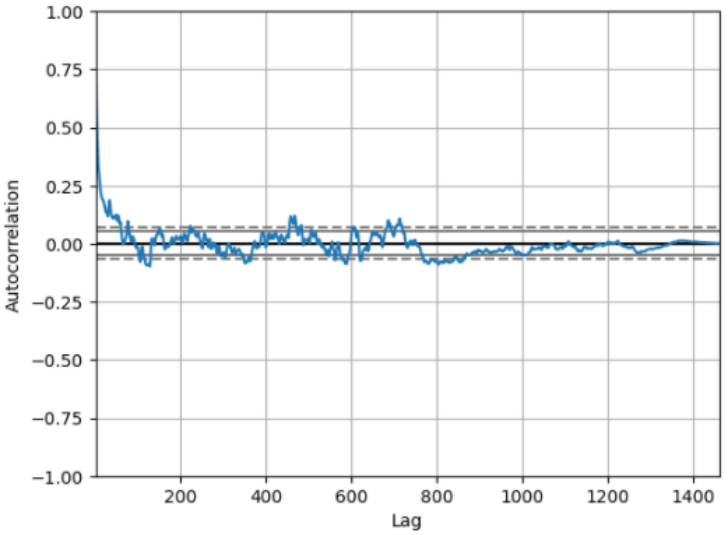


Figure 7: Autocorrelation Plot of O3 concentrations

between current observations and previous ones, during similar and opposite seasons of the year. This plot is a sign of low seasonality in the dataset [23].

Table 1: Statistics of the measured values during the study period

pollutants/Variables	Min	Max	Mean	Std Dev
O3 ($\mu g/m^3$)	0	210	33.58	23.06
Relative Humidity	-25	99	71.41	17.6
Temperature $°C$	-55	30	17.04	11.5
Wind Speed(ms)	0	3.8	0.99	0.70
Pressure(mmHg)	98	102	101	0.55
Precipation	0	76.4	2.13	6.16
Wind Direction	14	334	185.1	85.5
GWETTOP	0.2	0.95	0.63	0.15

If we want to find out more about the significance of the distribution, we can develop descriptive statistics for our time series which includes the average, standard deviation, minimum, and maximum of the data, view Table 1, we can deduce that the average of the distribution is equal to 33.58, that is a significant value and presents a sufficient reason to conduct for an exploratory study and then a predictive research as a way to safeguard our planet and individuals from this hazardous pollutant.

3.2 Process Of A Machine Learning Application

Preparation of Data

Information preparation is the procedure for transforming raw data into a format that is more appropriate for modeling. This initial phase of data processing is crucial in a machine learning application and might involve basic tasks such as feature engineering, data transformations, and data cleaning [24]. In addition, the data, as we might anticipate, influence the sorts of data processing that are selected.

Handling Missing Data

For a variety of causes, many real-world datasets could have missing values, which might dramatically lower the machine learning model's quality and have an impact on the outcomes. There are many ways to handle this problem in the current study; the K-Nearest Neighbor was used. In fact, KNN is more versatile overall than the other methods because it works for all kinds of data [25].

Scale Numerical Data Numerous machine learning methods worked better when the range of numerical input variables was regulated to a standard value. As a result, the tropospheric ozone (O3) dataset was directly submitted to the MinMaxScaler transform to normalize the input variables and scale values to an interval between 0 and 1. This approach succeeded excellently and achieved the desired outcomes [24]. Notably, when compared to other traditional alternatives, such as interpolation with the mean, the median, or the mode. Input and output attributes are not present in our time series data. Instead, we must select the variable to be forecasted and build all the input necessary to create predictions for the next time steps through feature engineering.

4 Results and Discussion

Artificial neural networks (ANNs) represent statistical techniques of machine learning that can uncover complicated links between inputs and outcomes. For time series data, other sequential data are best processed using recurrent neural networks (RNNs), a kind of neural networks. The work being done aims to construct two types of RNN: a simple RNN and LSTM models, realizing predictions of daily ozone (O3) in Tangier in order to be able to predict the emergence of pollution peaks and to take the necessary preventive actions. During operational forecasting, the two models will be applied to data on pollutants and weather conditions.

In our research, the first decision we make relates to how our dataset is split into the training dataset and the test dataset. After doing multiple tests, we found that 70 % for both models is the ideal percentage to use as the training database.

Time series data should be reframed as a supervised learning dataset, before starting to use machine learning algorithms. So the dataset will be separated into input as well as output samples following this formulation: we are going back as lookback timesteps (for 20 days), then we take the value directly after this time step as being the corresponding target.

In this stage, we built both a simple RNN and an LSTM model using only pollutant data of O3. Each of the RNN models requires the adjustment of a number of parameters, including: the number of layers, the number of neurons of each layer, and the activation function of neurons. In order to enable our LSTM to simulate non-linear interactions between inputs and outputs [3], we started with just one layer. And to train the network, we employed the descent of gradient technique.

Then, we tested three distinct data input configurations for information collected in Tangier city using consecutive: O3 concentration data, (O3 + NO2) pollution data, (O3 + NOX), and finally (O3 + meteorological) pollution and meteorological data. We repeatedly changed our

Table 2: Error indices

Model with input datasets	MSE	RMSE	MAE	d
SimpleRNN(O3)	0.59	0.77	0.46	0.84
SimpleRNN(O3+NO2)	0.56	0.74	0.46	0.85
SimpleRNN(O3+NOX)	0.54	0.73	0.44	0.86
SimpleRNN(O3+ METEO)	0.55	0.74	0.45	0.86
LSTM(O3)	0.55	0.74	0.45	0.86
LSTM(O3+NO2)	0.54	0.73	0.46	0.870
LSTM(O3+NOX)	0.52	0.72	0.45	0.877
LSTM(O3+ METEO)	0.61	0.79	0.49	0.84

configuration and investigated the previous input data configurations for data recorded and observed if that increased the performance. This process was repeated until the model was good and the configuration was kept with the best results. Within the training phase, efficiency was assessed using the Mean Squared Error (MSE), and during the test phase, the Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Index of Agreement (d) were determined.

In Table 2, the results of simple RNN and LSTM models are presented using various inputs. To examine the correlation between predictors and performance, we run each RNN experiment multiple times and provided the average values of error indices. Globally, our research results have shown that the best choice is to use a minimum number like two of hidden layers, and 64 and 32 neurons for each one, respectively. When it comes to the activation function, it is a function that transforms the signal entering (neuron) into an output signal (response). Generally, the most typical activation functions are the hyperbolic tangent and sigmoid logistic. After testing these two procedures, the results were encouraging and the sigmoid function was revealed to be the best performing one. Additionally, to ensure that the network can generalize when using novel data, it is essential to prevent over-fitting within the training process, For that reason, we applied the dropout method.

According to the numerical results shown in Table 2: In the first simpleRNN model, we can deduce that the error MSE has decreased and the agreement index d has increased by adding other predictors as follows: ($MSE = 0.59$ for O3), ($MSE = 0.56$ for (O3 + NO2)), ($MSE = 0.54$ for (O3 + NOX)), ($d = 0.84$ for O3), ($d = 0.85$ for (O3 + NO2)), ($d = 0.86$ for (O3 + NOX)).

For the second LSTM model, we can determine that the error MSE has been reduced and the agreement index d has increased by introducing more predictors in this order: ($MSE = 0.55$ for O3), ($MSE = 0.54$ for (O3 + NO2)), ($MSE = 0.52$ for (O3 + NOX)), ($d = 0.86$ for O3), ($d = 0.870$ for (O3 + NO2)), ($d = 0.877$ for (O3 + NOX)).

Finally, the effectiveness of the model is increased by including: NO2, NOX and meteorological factors ($d = 0.86$ for Simple RNN), ($d = 0.877$ for LSTM), ($MSE = 0.59$ for Simple RNN) and ($MSE = 0.52$ for LSTM) , which means that these factors have an important effect on the distribution of O3 and promote the pollution due to tropospheric ozone. The results also demonstrate that both algorithms worked well, but the LSTM model shows that it is more effective at forecasting O3 concentrations than the simple RNN architecture, especially at peaks levels.

In this step, we applied the LSTM model for a new period, like three months of 2014 to be able to compare them with our validation data in the same period. In Figure 9, we can observe that the predictions given by LSTM (yellow curve) are very close to the validated data, which are green. This means that the model is validated and both algorithms worked well, but the LSTM model shows that it is more effective at forecasting O3 concentrations than a simple RNN architecture.

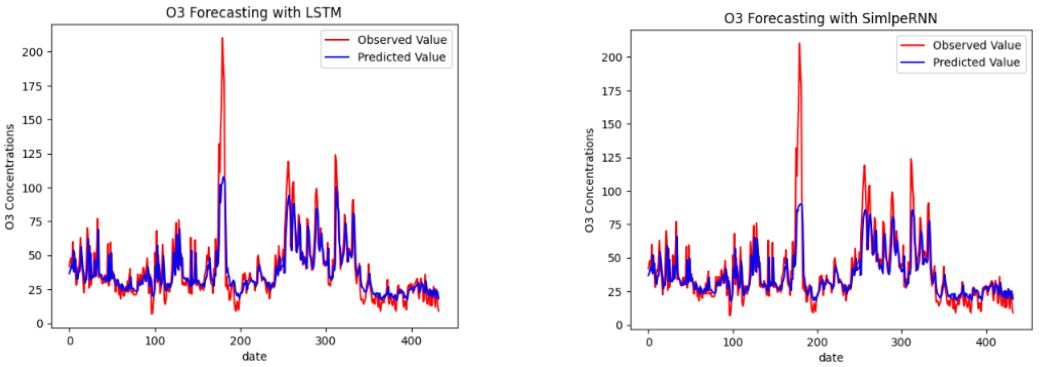


Figure 8: Graphical representation of observed and predicted values of the O3 time series for the two models

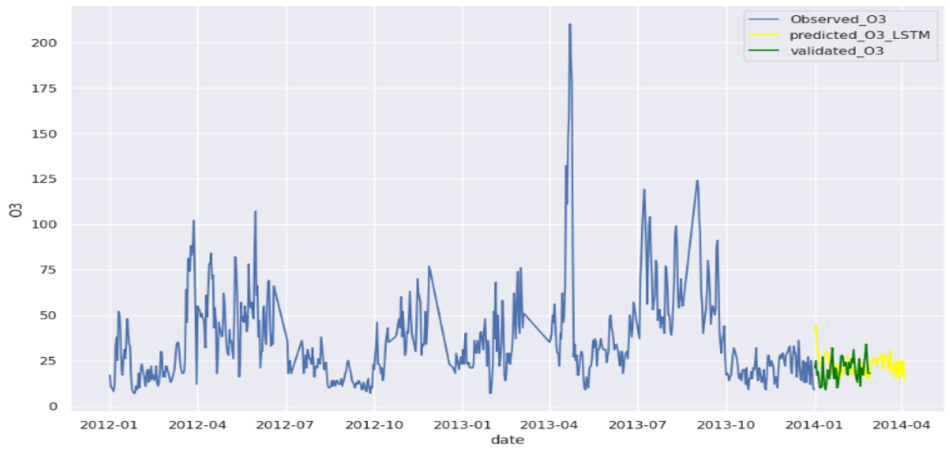


Figure 9: Forecasting 90 days of O3 concentration

5 Conclusion

In this work, we focused on predicting the ozone that caused daily air pollution levels in the city of Tangier the following day to avoid adverse effects from exposure and determine the time of peaks. The National Direction of Meteorology provided the data needed for the present research, which shows the number of daily concentrations of the O3 air pollutant from January 2010 to April 2014. Due to this, the Simple RNN and LSTM algorithms, two types of recurrent neural networks, were created, trained using pollution and meteorological data, and showed favorable findings. Its performance is improved by the introduction of NO2, NOX and meteorological variables, indicating that these factors have an effect on the distribution of O3 and promote pollution due to tropospheric ozone. The measured and forcasted concentrations of the O3 pollutant in Tangier also showed a good degree of agreement, for instance ($d = 0.86$ and $MSE = 0.54$ for SimpleRNN) and ($d = 0.87$ and $MSE = 0.52$ for LSTM) which demonstrated that the LSTM approach outperformed the Simple RNN methodology.

References

[1] P. R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H. O. Pörtner, D. Roberts, P. Zhai, R. Slade, S. Connors, R. Van Diemen, et al., Foreword technical and preface, *Climate Change and Land: An IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems.* **1**, 35–74 (2019).

[2] H. Boulaassal, S. Anaki, O. A. Yazidi, M. Maatouk, and M. Wahbi, Cartographie des changements de l'occupation du sol entre 2002 et 2016 à partir des images Landsat. Cas de la région Tanger Tetouan Al-Hoceima (Maroc), *African Journal on Land Policy and Geospatial Sciences,* **3(3)**, 14-31 (2020).

[3] W. Tamas, G. Notton, C. Paoli, C. Voyant, M.-L. Nivet, and A. Balu, Urban ozone concentration forecasting with artificial neural network in Corsica, *Modelling in Civil Environmental Engineering,* **10(1)**, 29–37 (2014).

[4] A. Haines, M. Amann, N. Borgford-Parnell, S. Leonard, J. Kuylenstierna, and D. Shindell, Short-lived climate pollutant mitigation and the Sustainable Development Goals, *Nature Climate Change* **7(12)**, 863–869 (2017).

[5] C. C. Lim, R. B. Hayes, J. Ahn, Y. Shao, D. T. Silverman, R. R. Jones, C. Garcia, M. L. Bell, and G. D. Thurston, Long-term exposure to ozone and cause-specific mortality risk in the United States, *American journal of respiratory and critical care medicine,* **200(8)**, 1022–1031 (2019).

[6] M. Weiss, P. Bonnel, J. Kühlwein, A. Provenza, U. Lambrecht, S. Alessandrini, M. Carriero, R. Colombo, F. Forni, G. Lanappe, et al, Will Euro 6 reduce the NOx emissions of new diesel cars?–Insights from on-road tests with Portable Emissions Measurement Systems (PEMS), *Atmospheric Environment* **62**, 657–665 (2012).

[7] W. W. Tamas, G. Notton, C. Paoli, M.-L. Nivet, and C. Voyant, Hybridization of air quality forecasting models using machine learning and clustering: An original approach to detect pollutant peaks, *Aerosol and Air Quality Research* **16(2)**, 405–416 (2016).

[8] G. P. Zhang and D. M. Kline, Quarterly time-series forecasting with neural networks, *IEEE transactions on neural networks*, **18(6)** , 1800–1814 (2007).

[9] J. Chanchaichujit, Q. C. Pham, and A. Tan, Sustainable supply chain management: A literature review of recent mathematical modelling approaches, *International Journal of Logistics Systems and Management*, **33(4)** , 467–496 (2019).

[10] N. Garg, K. Soni, T. Saxena, and S. Maji, Applications of Autoregressive integrated moving average (ARIMA) approach in time-series prediction of traffic noise pollution, *Noise Control Engineering Journal* **63(2)** , 182–194 (2015).

[11] Hong, Fugui, et al , Hourly ozone level prediction based on the characterization of its periodic behavior via deep learning, *Process Safety and Environmental Protection* **174** , 28–38 (2023).

[12] H. Oufdou, L. Bellanger, A. Bergam, A. El Ghaziri, K. Khomsi, E. M. Qannari, et al., Comparison of different regularized and shrinkage regression methods to predict daily tropospheric ozone concentration in the grand Casablanca area, *Advances in Pure Mathematics*, **8(10)** , 793 (2018).

[13] A. C. Comrie, Comparing neural networks and regression models for ozone forecasting, *Journal of the Air & Waste Management Association*, **47(6)**, 653–663 (1997).

[14] I. Alon, M. Qi, and R. J. Sadowski, Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods, *Journal of retailing and consumer services*, **8(3)**, 147–156 (2001).

[15] J. C. B. Gamboa, Deep learning for time-series analysis, *JarXiv preprint arXiv:* **1701.01887** , (2017).

[16] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, LSTM-based encoder-decoder for multi-sensor anomaly detection, *arXiv preprint arXiv:* **1607.00148** , (2016).

[17] L. A. Steffenel, V. Anabor, D. Kirsch Pinheiro, L. Guzman, G. Dornelles Bittencourt, and H. Bencherif, Forecasting upper atmospheric scalars advection using deep learning: an O 3 experiment, *Machine Learning*, **112(3)** , 765–788 (2023).

[18] J. Brownlee, *Machine learning mastery*, Machine Learning Mastery, (2022).

[19] J. Brownlee, *Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python*, Machine Learning Mastery, (2020).

[20] M. Bouaziz, *Réseaux de neurones récurrents pour la classification de séquences dans des flux audiovisuels parallèles*, Diss. Université d'Avignon, (2017).

[21] Y. Ensafi, S. H. Amin, G. Zhang, and B. Shah, Time-series forecasting of seasonal items sales using machine learning–A comparative analysis, *International Journal of Information Management Data Insights*, **2(1)**, 100058 (2022).

[22] C. J. Willmott, S. M. Robeson, and K. Matsuura, A refined index of model performance, *International Journal of climatology*, **32(13)**, 2088–2094, (2012).

[23] A. Nielsen, *A Practical time series analysis: Prediction with statistics and machine learning*, O'Reilly Media, (2019).

[24] T. C. Nokeri, *Data Science Solutions with Python*, Springer, (2021).

[25] A. Burkov, *The hundred-page machine learning book*, Andriy Burkov Quebec City, QC, Canada, **1** (2019).

Author information

Nisrine Marrakchi, SMAD, FPL, Abdelmalek Essaadi University, Tetouan, Morocco.
E-mail: nisrine.marrakchi@etu.uae.ac.ma

Amal Bergam, SMAD, FPL, Abdelmalek Essaadi University, Tetouan, Morocco.
E-mail: bergam@uae.ac.ma

Hanane Fakhouri, SMAD, FPL, Abdelmalek Essaadi University, Tetouan, Morocco.
E-mail: fakhouri@uae.ac.ma

Kenza Khomsi, DGM, National Climate Center, Air Quality Department,General Directorate of Meteorology, Morocco.
E-mail: k.khomsi@gmail.com