ORIGINAL ARTICLE

Long short-term memory artificial neural network approach to forecast meteorology and PM$_{2.5}$ local variables in Bogotá, Colombia

Alejandro Casallas[1,2] · Camilo Ferro[1,5] · Nathalia Celis[3] · Marco Andrés Guevara-Luna[4,5,6,7] · Caroline Mogollón-Sotelo[1] · Fredy Alejandro Guevara-Luna[4,5] · Martha Merchán[8]

Abstract

In places with complex topography, the reproduction of atmospheric dynamics is challenging and resource demanding. Recently, machine learning has been successfully implemented to forecast pollution and weather variables. LSTM (long short-term memory) networks have the potential to improve the forecasting precision on different theoretical fields. Despite this advantage, they have not been widely used in the tropics for this purpose. This research aims to implement a LSTM to forecast PM$_{2.5}$ and meteorological variables in a tropical mountainous city. The model was trained with 7 years of data from the local air quality monitoring network. The implemented model forecasts 42 days, evaluated using statistical scores and benchmarks. More than 95% of PM$_{2.5}$ values, radiation (99%), air temperature (98%), relative humidity (95%), wind speed (94%), and the u-component (91%) have excellent or good benchmarks. The v-component and the wind direction got the lowest percentage of excellent or good values (50%). We compared our results with other models that have focused on forecasting these variables in similar places and observed that the LSTM approach results are the best, especially for PM$_{2.5}$ and wind direction. We found its accuracy can be affected by rapid changes in the tendency of the data that do not occur as a consequence of the diurnal tide. The LSTM model was validated as a tool to predict meteorological variables and PM$_{2.5}$ (24 h in advance) in a tropical mountainous city and can be used as a reliable input in air quality early alert systems.

Keywords Weather · Air quality alert system · PM$_{2.5}$ · LSTM · Complex terrain

✉ Alejandro Casallas
  acasallas@ictp.it

1   School of Exact Sciences and Engineering (ECEI), Sergio Arboleda University, Bogotá 111071, Colombia

2   Earth System Physics, Abdus Salam International Centre for Theoretical Physics, 34151 Trieste, Italy

3   Institute of Environmental Studies and Services (IDEASA), Sergio Arboleda University, Bogotá 111071, Colombia

4   Grupo de Investigación de Calidad del Aire (GICA), Chemical and Environmental Engineering Department, Universidad Nacional de Colombia, Bogotá 111071, Colombia

5   Scientific Research Department, Smart and Simple Engineering S.A.S (S&S S.A.S), Bogotá 111071, Colombia

6   Conservación, Bioprospección y Desarrollo Sostenible (COBIDES), Universidad Nacional Abierta y a Distancia, Bogotá 111071, Colombia

7   LIVE-Laboratoire Image Ville Environnement, Université de Strasbourg, 3 rue de l'Argonne, 67000 Strasbourg, France

8   Faculty of Education, Research Group Conciencia, Antonio Nariño University, Bogotá 111071, Colombia

Introduction

Air quality and meteorology are important for risk assessment. A decrease in air quality indexes is related to economic, industrial, and population growth in large cities such as Bogotá (e.g., Rojas 2004; Sokhi et al. 2021). Among the criteria pollutants, the 2.5-μm particulate matter (PM$_{2.5}$) has a high impact on public health, as it has been deemed as especially harmful to human health and has a great presence and persistence in cities (Gehling and Dellinger 2013; Ostro and Chestnut 1998). Likewise, the behavior of meteorological variables can produce landslides, floods and infrastructure damages, decrease energy production and even affect food security (UNDRR 2005), which may lead to high government costs related to hospitals, roads, energy and infrastructure.

To minimize the risk and economic impact generated by meteorological variables and pollution, researchers and countries have mainly used and evaluated air quality models (AQM) (e.g., Weather Research and Forecasting model

coupled with Chemistry—WRF CHEM, Copernicus Atmosphere Monitoring Service—CAMS) and machine learning models (e.g., support vector machine—SVM; artificial neural network—ANN; long short-term memory—LSTM). The first kind of model applies atmospheric principles (physical and chemical) to forecast air quality and meteorological variables. However, these systems are computationally expensive, and for low latitudes they tend to have uncertainties associated with the data availability, the high importance of local land and topography properties which could largely change the dynamic and thermodynamic characteristics of an area (e.g., Holton 2004; Adams et al. 2013; Freitag et al. 2018). For example, in an equatorial city like Bogotá, Colombia, convection principally depends on local moist thermodynamics (Casallas et al. 2021), which the coarse resolution of the global circulation models ($\approx$ 10 km) cannot fully resolve. The second kind of model is an alternative to deterministic and statistical methods of prediction and uses different techniques to represent the complex behavior of air pollutants or atmospheric variables. In the field of machine learning, a new approach has delivered good results: the LSTM, a specific recurrent neural network (RNN) architecture designed to simulate temporal sequences and their long-range dependencies (Hochreiter and Schmidhuber 1997a, b). This approach has been evaluated by various authors (e.g., Feng et al. 2019; Qing and Niu 2018; Baker and Foley 2011) to simulate atmospheric variables and $PM_{2.5}$, proving capable of producing good statistical parameter results.

AQM like the WRF-Chem has been widely implemented around the world. Some studies (e.g., Tuccella et al. 2012; Vera-Vela et al. 2016) have evaluated this model, finding it capable of forecasting weather variables and pollutant concentrations on an hourly basis. Additionally, the model may determine some of the most important properties of fine particles in the area (i.e., such as mass size distribution and chemical composition). Different studies have validated physical models using statistical parameters that evaluate the atmosphere and pollutant dynamics in Colombia (e.g., Kumar et al. 2016; González et al. 2018; Barten et al. 2019; Casallas et al. 2020; Guevara-Luna et al. 2020). Nevertheless, physical models are computationally expensive compared to empirical models. For this reason, machine learning tools with high precision have been implemented and are currently an interesting approach to simulate and forecast air atmospheric variables (e.g., Zhou et al. 2019; Mogollón-Sotelo et al. 2020).

It is important to highlight that machine learning has achieved remarkable results when simulating complex problems with multiple elements, as in the case of atmospheric modeling. The main available machine learning models are SVMs, ANNs, and LSTMs. SVMs, for instance, have been used (e.g., Steinwart and Christmann 2008; Westerlund et al. 2014; Martinez et al. 2018; Murillo-Escobar et al. 2019;

Mogollón-Sotelo et al. 2020) to simulate meteorological variables and $PM_{2.5}$ concentrations and, after comparing the simulations with observations, they have been acknowledged as very precise when simulating atmospheric variables in complex terrains and in places without mountain ranges. However, SVMs have some limitations such as error accumulation and forecasting errors when data have sudden changes in their tendency (Mogollón-Sotelo et al. 2020). ANNs, on the other hand, are a computer model capable of learning the distribution of the data from its sample. Some studies used ANNs to simulate the behavior of particulate matter (PM_{10} and $PM_{2.5}$), wind, temperature and precipitation (e.g., Erdil and Arcaklıoğlu 2013; Pérez and Gramsch 2016; Wang and Sun 2019; Zhao et al. 2019) and concluded that this approach has a good performance when simulating meteorological variables. Franceschi et al. (2018), for example, used an ANN to simulate particulate matter in Bogotá-Colombia and trained the network with meteorological variables, to forecast them. LSTM networks, in turn, are a type of ANN capable of saving information in batches, which improves their precision (i.e., they require more computational power than a normal neural network, but deliver better results). LSTMs have been used to predict meteorological variables such as solar irradiance, temperature, relative humidity, wind speed and precipitation with high accuracy (e.g., Maqsood et al. 2014; Qing and Niu 2018; Karevan and Suykens 2020). Other studies (e.g., Tong et al. 2019) have used this type of network to forecast $PM_{2.5}$, reporting its capability of predicting pollutant concentration trends precisely. Thus, this type of network has been proven to work accurately when predicting atmospheric variables and pollutants, having the potential to be applied in tropical cities with high altitude and complex terrain (an extremely complex system).

Despite the proven success of LSTMs when predicting PM concentration and meteorological variables, they have not been widely used in tropical areas with complex topography. This research aims to implement an LSTM network to forecast $PM_{2.5}$ concentrations and meteorological variables (temperature, radiation, humidity, wind speed), with an additional new approach to simulate wind fields (i.e., using u and ν components), in a tropical city with a complex topography (Bogotá-Colombia), using data inputs from the Air Quality Monitoring Network of Bogotá (AQMNB).

Method

Study area and case study

Bogotá, Colombia (4.60971, − 74.08175 WGS84) is a large city with a population of approximately 8 million inhabitants. The city has an area of almost 1800 km², distributed

into 25% of urban and 75% of rural areas. The city has a complex topography, due to its location in the eastern slope of the Andes Mountain range at 2550–2620 m.a.s.l.

The study area has an AQMNB that comprises 12 stations along the urban area of the city, separated by localities (Fig. 1) (SDA 2013). This network measures criteria pollutants and meteorological variables such as solar radiation, temperature, precipitation, relative humidity, wind speed, and wind direction. We use 12 air quality stations with at least 75% of available data for every variable mentioned above, in the period between 2014 and 2019. The majority (11/12) of the stations have more than 88% of the data available, and the S7 station has 76% due to sensor maintenance. Thus, the region and its air quality network are suitable for the analysis of the performance of LSTMs regarding meteorological variables and PM$_{2.5}$ behavior in a tropical city with complex terrain, and its subsequent impacts on risk assessment. For this reason, the LSTM model was used in Bogotá during the days with an *unhealthy* qualification according to the International Air Quality Index (AQI) in 2019 (17 days in total) [the index defined the level of population risk as detailed by EPA (2014)]. 65% of these days occurred in the local dry season, while 35% occurred in the local wet season. To have more robust results, we simulated 25 random days, so that the model forecasting capabilities can be assessed regarding days with low, mid and high pollution levels, even though these findings are not shown here, since they have a very similar behavior to the 17 highly polluted days.

Model description

ANNs are a set of algorithms inspired by the process of human learning, which recognize numerical patterns contained in vectors that can be translated into images, sound, text or time series (Bengio et al. 1994; Grivas and Chaloulakou 2006). ANNs can adapt to changes in the input data since they are generated through prior training based on cause–effect data sets. Thereafter, the model can be used to generate predictions of variables linked to dynamic systems. This feature gives it a valuable forecasting capability, useful for atmospheric dynamics research and air quality study.

An ANN contains layers of interconnected nodes. Each node is a perceptron that works similarly to multiple linear regression. The perceptron reads the input values; it then adds them according to settled weights and inserts the result into a trigger function that generates the final result. In a multilayer ANN, the perceptions are arranged in a set of interconnected layers: internal layers and external layers (input and output layers). The internal layers perform
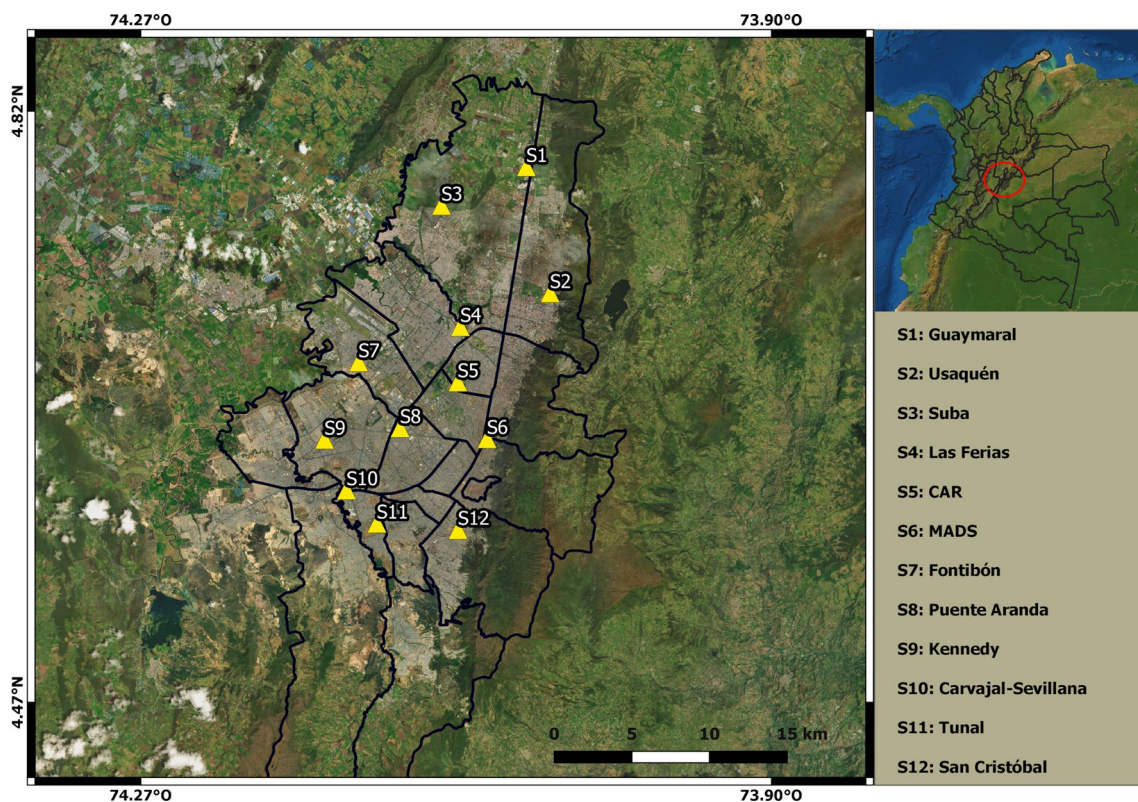
Fig. 1 Geographic location of the 12 surface monitoring stations used in the city of Bogotá-Colombia. The red circle signals the location of the city within the country

non-linear transformations of the inputs, adjusting the weighting factors for the data until the margin of error is reduced to its minimum. This type of model assumes that the internal layers extrapolate the main characteristics of the input data, providing them with significant predictive potential.

LSTMs are one of the most commonly used RNNs. This type of network can learn long-term dependencies using three gates to add, remove, and update the hidden state (memory). Every gate has three elements, a neuron, a sigmoid function, and a multiplier element; these tools decide the weights and activation in the network (Hochreiter and Schmidhuber 1997a, b).

One important tool in the LSTM is the *forget gate* layer, used to decide the information that will be discarded and not taken to the state cell, Eq. (1), where h_{t-1} is the previous hidden state, x_t is the current input, t is the index for the timestep, W_f and U_f contain, respectively, the weights of the input and recurrent connections, W_f, U_f, and b_f are the coefficients learned during training, σ is the sigmoid activation function and f_t is the output vector. If any of the values in the output vector (f_t) is zero, or close to zero, the information will be discarded. In contrast, if it reaches values equal or close to one, the information is saved and reaches the state cell (Staudemeyer and Morris 2019).

$$f_t = \sigma\left(W_f \times h_{t-1} + U_f \times x_t + b_f\right). \tag{1}$$

The input gate allows the cell memory to update. To achieve this, a sigmoid layer chooses the values that will be preserved through Eq. (2), where h_{t-1}, x_t, W_i, U_i, b_i, σ are the same for Eq. (1) and u_t is the output vector. Values close to one will be preserved and those close to zero will be updated. Then, a layer with a *tanh* activation creates a vector of new candidate values as in Eq. (3), where W_c and b_c are learned coefficients during the training stage, tan$h$ is an activation function and v_t are the candidate values. Thereafter, the vector with the updated values is created with Eq. (4), where f_t is the output vector of the forget gate, C_{t-1} is the state value of the last cell, u_t is the vector of values to be conserved, v_t are the candidate values and C_t is the vector of the updated cell values (Staudemeyer and Morris 2019; Zhao et al. 2019):

$$u_t = \sigma\left(W_i \times h_{t-1} + U_i \times x_t + b_i\right), \tag{2}$$

$$v_t = \tanh\left(W_c \times h_{t-1} + U_c \times x_t + b_c\right), \tag{3}$$

$$C_t = f_t \times C_{t-1} + u_t \times v_t. \tag{4}$$

The output gate decides which values are generated as filtered outputs from the hidden layers. To achieve this, a sigmoid layer decides which parts of the cell state have to

be generated through Eq. (5) where o_t is the vector with the filtered values, σ is the sigmoidal function, h_{t-1} is the last hidden state, x_t is the current input, and W_0, U_0, and b_0 are learned coefficients during the training stage. Then, a layer with a tan$h$ function is used to weigh the output values, deciding their level of importance (i.e., varying from − 1 to 1). Thereafter, this must be multiplied by the sigmoid output as shown in Eq. (6), where h_t is the new hidden value, C_t is the vector with the updated cell values and o_t is the vector with filtered values (Staudemeyer and Morris 2019):

$$o_t = \sigma\left(W_o \times h_{t-1} + U_c \times x_t + b_0\right), \tag{5}$$

$$h_t = o_t \times \tanh\left(C_t\right). \tag{6}$$

This model aims to forecast $PM_{2.5}$ concentrations and meteorological variables, with an hourly time resolution and 24 h in advance. Figure 2 shows the flow of information for the training and prediction stages in the model developed for this research.

Figure 2 summarizes the process of development and use. The first step consists in the surface data retrieval and these data were downloaded for the time period between January 2014 and December 2019 from the AQMNB: hourly data $PM_{2.5}$ (μg m^{-3}), relative humidity (%), temperature (°C), radiation (W m^{-2}), wind speed (m s^{-1}), and wind direction (°). Within these data, it is possible to find missing values, which were imputed using the hourly mean of the data, e.g., a 13 h nan value of station S1 (any day) is replaced by the 13 h mean value of S1 station for the period of 2014–2019 similar to that by Mogollon-Sotelo et al. (2020). The input data normalization uses a technique often applied to prepare data for automatic learning. It aims to change the values of the numeric columns in the data set for its use at a common scale, improving the models' numerical stability and reducing computational time (Önskog et al. 2011). Once with the normalized data, we arrange the training sets since the input training X-vector contains 48 values (observations), necessary for the prediction. The output training Y-vector contains 24 elements (forecast) [to simulate the direction, the wind is divided into components (*u* and *v*)]. The model evaluation aims to assess the results obtained by the ANN, and its precision when it reproduces the variables trends are evaluated with a loss indicator. Finally, for the Neural Network Model definition we used the *Keras* library. This tool uses the *Sequential* class as its main structure, which allows for the use of simple layers. Each layer has exactly one input tensor and one output tensor. In this step, the input number and output nodes, layer types, number of layers, number of neurons, training epochs, activation function, loss function, and batch size are defined. This evaluated model is then ready to be used for forecasting (Fig. 2).

Fig. 2 Flow of information in the model developed based on the LSTM
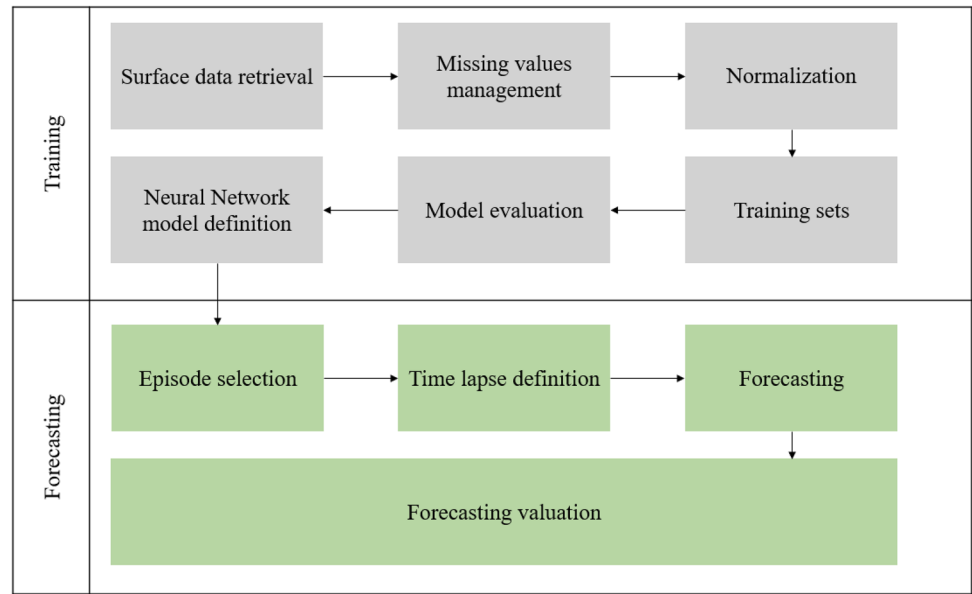
The forecasting stage begins with an episode selection, in which the time period of prediction is fixed for the model to make the prediction based on the trained ANN. A time lapse is created to predict concentrations 48 h before the episode. Then, the trained neural network is charged and predictions are generated. Finally, the quality of the predictions is evaluated with statistical indicators (the wind components (u and v) are predicted, evaluated, and then converted into direction).

Setup

The model comprises an input layer (X-vector), defined by the number of observations necessary to make a prediction (48 h), three hidden layers (2 LSTM and 1 Dense), and an output layer (Y-vector), which has a vector of 24 elements (forecast). To find the best activation function, number of neurons and optimizer, we first based the model on Huang and Kuo (2018). Then the grid search technique (Ndiaye et al. 2019) is performed using 216 possible combinations, in which changes in the batch size (e.g., 6, 12, 24, 32), the number of neurons, network structure, activation functions (e.g., relu, tanh, sigmoid) and optimizers (e.g., Adam, RMSprop, Adadelta) are performed. The best combination of hyperparameters found is presented in Table 1 and it is fairly similar to the one found by Huang and Kuo (2018) for their LSTM network. For the training task, to avoid overfitting, we used a dropout function after the first LSTM layer and an early stopping method (e.g., Prechelt, 1998). The dropout function randomly deactivated 20% (i.e., previously defined in the grid search step) of the neurons.

Table 1 Hyperparameters used to set up the long short-term memory (LSTM) artificial neural network (ANN) implemented to forecast $PM_{2.5}$, radiation, temperature, relative humidity, wind speed, and wind direction

Parameter	Value
Epochs	150
Batch size	12 values
Activation function	tan$h$, tan$h$, linear, linear[a]
Neurons per layer	96, 72, 48, 24[a]
Optimizers	Adam
Loss	Root mean square error (RMSE)

[a]Parameter for each of the used layers (LSTM–LSTM–dense–dense)

The training process for the 12 stations took 23 h using hardware consisting of 16 GB of RAM and one 8-core processor. The environment was configured on the operating system Linux Ubuntu V 18.04. Programming language Python 3.6 with the packages Tensorflow V 1.13.1 and Keras V 2.3.1 was used.

Model evaluation

The quantitative evaluation of the model was based on three statistical parameters: index of agreement (IOA), root mean square error (RMSE), and the correlation coefficient (r). The classification criterion for each parameter (Table 2) has been established by US EPA (2000), Emery et al. (2001), Boylan and Russell (2006), and Guevara-Luna et al. (2020). However, in this study, the criterion was modified to improve its rigorousness. The mean gross error (MGE) and the mean absolute bias (MAB) were used to evaluate the wind direction, as recommended by McNally (2009). Notice that the

Table 2 Statistic parameters and benchmarks defined to evaluate the model for PM$_{2.5}$ (μg m^{-3}), air temperature (°C), relative humidity (%), solar radiation (W m^{-2}), wind speed, u and v components (m s^{-1}), and wind direction (°)

Variable	Evaluation parameter	Benchmarks		
		Excellent	Good	Bad
PM$_{2.5}$	R	≥ 0.9	(0.9–0.5]	< 0.5
	RMSE (μg m^{-3})	≤ 10	(10–15]	> 15
	IOA	≥ 0.9	(0.9–0.6]	< 0.6
Air temperature (2 m)	R	≥ 0.9	(0.9–0.7]	< 0.7
	RMSE (°C)	≤ 1	(1–3]	> 3
	IOA	≥ 0.9	(0.9–0.6]	< 0.6
Relative humidity (RH)	R	≥ 0.9	(0.9–0.7]	< 0.7
	RMSE (%)	≤ 5	(5–10]	> 10
	IOA	≥ 0.8	(0.8–0.6]	< 0.6
Solar radiation	R	≥ 0.9	(0.9–0.7]	< 0.7
	RMSE (W m^{-2})	≤ 50	(50–100]	> 100
	IOA	≥ 0.9	(0.9–0.6]	< 0.6
Wind speed (10 m)	R	≥ 0.7	(0.7–0.5]	< 0.5
	RMSE (m s^{-1})	≤ 0.5	(0.5–1]	> 1
	IOA	≥ 0.7	(0.7–0.4]	< 0.4
U-component	R	≥ 0.7	(0.7–0.5]	< 0.5
	RMSE (m s^{-1})	≤ 0.8	(0.8–1.5]	> 1.5
	IOA	≥ 0.7	(0.7–0.4]	< 0.4
V-component	R	≥ 0.7	(0.7–0.5]	< 0.5
	RMSE (m s^{-1})	≤ 0.8	(0.8–1.5]	> 1.5
	IOA	≥ 0.7	(0.7–0.4]	< 0.4
Wind direction (10 m)	MGE (°)	≤ 35	(35—60]	≥ 60
	MAB (°)	≤ 15	(15—30]	≥ 30

RMSE, MGE and MAB are the only statistical parameters that have units (depending on each variable), since the other parameters are indices.

Results and discussion

Temporal dimension

The LSTMs forecasting results for meteorological variables and PM$_{2.5}$ concentrations were compared with the observed data using the hourly mean value of every station and simulated day (Fig. 3). The results show high precision when forecasting the studied variables: radiation, temperature, relative humidity, and wind speed. On the one hand, small changes in the diurnal cycle and non-rapid or steeper changes in the tendency of the data are evidence of the LSTM's high performance. On the other hand, wind direction before noon and between 14 and 18 h has short and fast changes that the model represents with a good correlation but with an underestimation of the values. The model can accurately represent PM$_{2.5}$ concentration since it is capable of representing the peak concentration (i.e., 9 h) and its subsequent reduction. In the afternoon, the performance of the simulation is slightly reduced due to

some changes and noise in the data, which the model is not able to completely identify. In spite of this, the general trend of the variables can be reproduced using this forecasting approach.

Figures 4 and 5 show two specific stations, which are the only stations that measure all variables under study. The behavior of the variables in S9 and S12 supports the fact that the model has limitations when the data has rapid changes in its tendency. Figure 5 shows that wind direction and wind speed are not as well reproduced by the model as the other variables. The observed wind field variables have steeper changes in the tendency that are not well represented because the LSTM tries to follow the diurnal tide of the variable. The model's accuracy is reduced when rapid changes in the tendency of the data are presented. This problem does not occur in regard to the other variables.

The LSTM represents the peaks of PM$_{2.5}$ in S9. These stations have two peaks, throughout the day, identified by the model. However, the highest values are underestimated. The remaining variables have similar behavior to those presented in Fig. 3, suggesting that LSTM simulations have great accuracy when the variables have smooth and slow changes in their tendency, but also showing that its accuracy decreases for variables with steeper changes (wind direction) (as also seen in the supplementary material Figs S2 to S11).
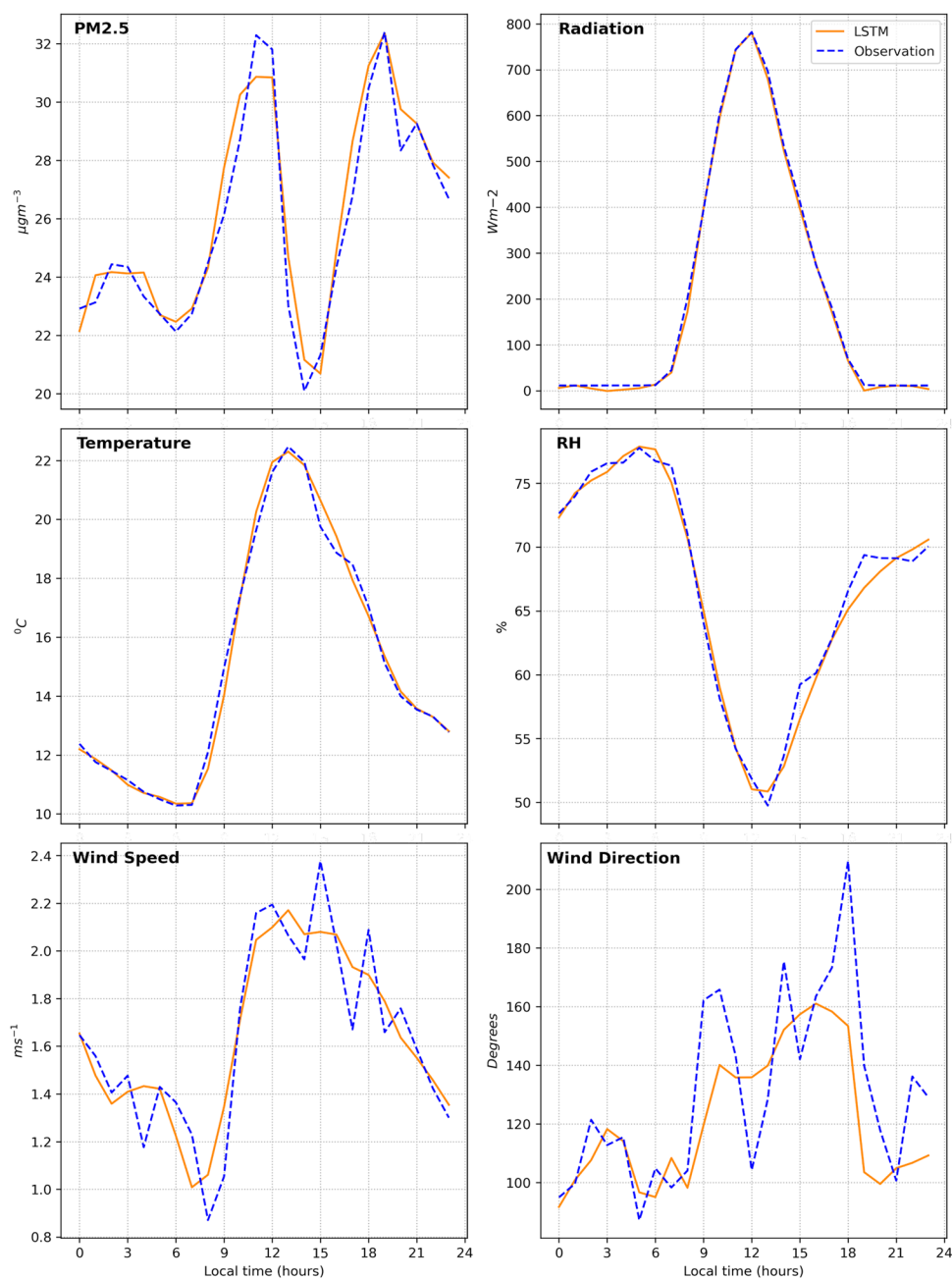
Fig. 3 Time series of the modeling results using LSTM and observed data. Composite of all monitoring stations during the 17 highly polluted simulated days. Variables: PM2.5 concentration (µg m^{-3}), radiation (W m^{-2}), air temperature (°C), relative humidity (%), wind speed (m s^{-1}) and wind direction (°)
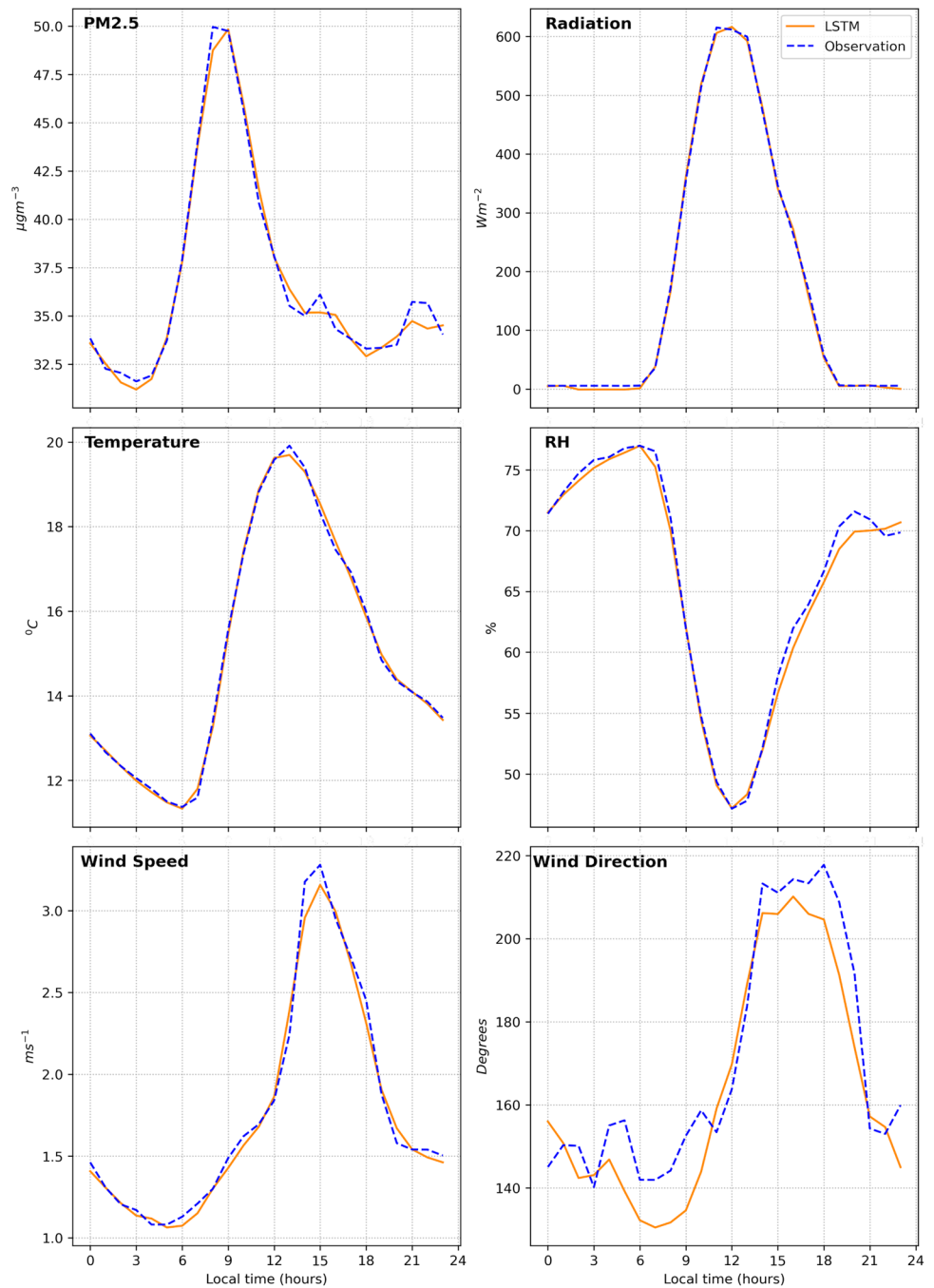
A daily evaluation of all the simulated days in every station (not shown here) shows that the model fairly follows the diurnal cycle of all the variables. The model tends to produce errors to represent variables that do not have a clear trend, which means that the behavior and magnitude of variables such as air temperature, relative humidity, wind speed, u-component, and radiation are very well represented for all the 42 days. On the other hand, the variables with steeper changes in their trends can produce large uncertainties, the PM$_{2.5}$, the wind direction and the v-component are examples

of this. The LSTM fails to represent higher changes in the magnitude of the PM$_{2.5}$, especially in stations that normally do not have large values or changes in the pollutant (e.g., S7 and S8). In terms of the wind direction, it is directly related to the v-component; the v-component has steeper changes around the day. These changes do not have trends or tendencies that the model can learn, so it fails to describe the magnitude and behavior of the variable in 23 of the 42 days evaluated (not show). It is possible that this issue could be solved following the novel approach of Sayeed et al. (2021),

Fig. 4 Time series of the modeling results using LSTM and observed data. Composite of the S12 station during the 17 highly polluted simulated days. Variables: PM$_{2.5}$ concentration (μg m^{-3}), radiation (W m^{-2}), air temperature (°C), relative humidity (%), wind speed (m s^{-1}) and wind direction (°)

in which they couple the WRF model with a neural network, or with the convolutional network (CNN) merged with a bidirectional LSTM (BDLM), since a CNN is able to select the important features of the data to learn, and the BDLM can improve the learning via its nowcasting ability.

Model evaluation

To evaluate the statistical accuracy of the model, we used the evaluation parameters and benchmarks in Table 2. Correlation (*r*), RMSE and IOA were calculated separately for the 42 simulated days (Figs. 6, 7). Wind direction for the MAB

and MGE were computed (Fig. 8) (to prevent errors caused by the mean value of the wind direction).

For PM$_{2.5}$ (Fig. 6), 63.24% of the modeled data have a good correlation with the observed data and 35.29% have an excellent correlation. The remaining values have correlations below this range; thus, they are considered deficient. The mean correlation for this variable is 0.99 (excellent). Regarding the RMSE, the models' results have a percentage of 94.12% for excellent values and 4.90% for good values, with a mean value of 0.58 μg m^{-3} (excellent). Finally, the IOA parameter delivers 66.67% of the modeled days with

Fig. 5 Time series of the modeling results using LSTM and observed data. Composite of the S9 station during the 17 highly polluted simulated days. Variables: PM2.5 concentration ($\mu g\ m^{-3}$), radiation ($W\ m^{-2}$), air temperature (°C), relative humidity (%), wind speed ($m\ s^{-1}$) and wind direction (°)
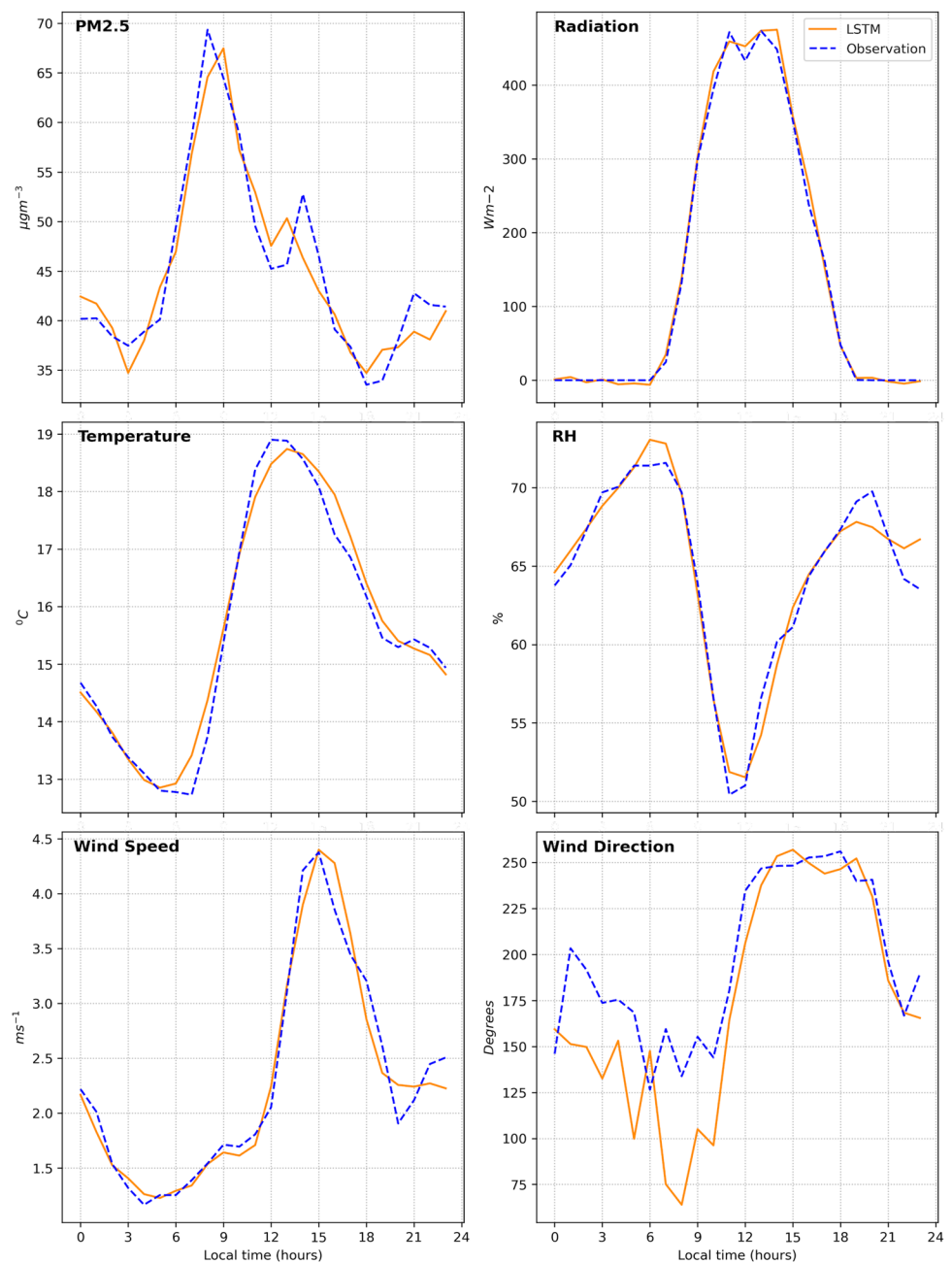
excellent predictions, 33.33% with good values, and no deficient values.

The meteorological variables (Fig. 4) reported excellent values for the (r) parameter: between 83.19 and 96.08%, except for the v-component, with only 13.9%. The percentages of values cataloged as bad are low for all the variables studied (between 0 and 8.56%), except for the v-component, which presents the highest percentage: 66.31%. The average correlation values are in a range of 0.994 and 0.999.

For the RMSE evaluation parameter, the mean values are in a range between 0.08 and 5.58 $\mu g\ m^{-3}$, and the percentages range from 45.45 to 7.49%. The variable with

the highest percentage of excellent values is temperature (91.44%) and the lowest percentage belongs to the v-component (20.86%). Finally, no variable reports bad IOA values, and the best results are for radiation, with 100% of the results classified as excellent. The variable with the worst results is the v-component, which reports 59.36% of excellent values and the remaining percentage stays within the "good" range.

The values obtained from the statistical evaluation of wind direction for MAB (in degrees) and MGE (in degrees) are relatively low, with values between 35.29 and 58.82%, respectively, in comparison to the performance of the model when predicting the other studied variables. The percentage
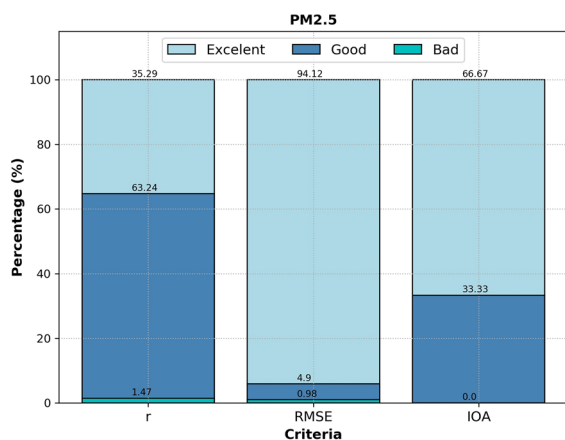
Fig. 6 Model's performance in terms of the evaluation parameters and benchmark distribution for all 12 stations and the 42 days: correlation (*r*), RMSE and IOA for the variable of PM$_{2.5}$

of good values obtained with the model are 27.81% for MAB and 31.55% for MGE and the remaining percentages represent excellent values. In addition, the mean MAB and MGE for this variable are 6.28 and 9.26 (excellent).

Forecasting analysis

Currently, LSTMs are not completely understood. We are able to describe how they work (i.e., the mathematics behind inverse propagation or optimization theorems), but we do not know how they learn (knowledge extraction) or the nature of certain errors (e.g., d'Avila Garcez et al. 2001; Abiodun et al. 2018; Remm and Alexandre 2002). In this study, we have shown some of the characteristics of the LSTM and the way the weights follow the diurnal cycle (hourly mean) of the data. To achieve this, data examination was conducted (i.e., calculating the interquartile range shown in Table 3), leading to the observation of the limitations (i.e., the model is not able to represent the behavior of variables that have very rapid changes in their tendency) and characteristics of the implemented LSTM model.

The interquartile range shows the fluctuation between the data in each of the time series forecasted for each of the studied variables. The highest values relate to radiation and wind direction. Regarding the first variable, this phenomenon may be caused by the abrupt changes inherent to its diurnal trend. However, since radiation daytime cycles are very similar, the LSTM is capable of learning this behavior. On the other hand, wind direction has the highest variability of all the parameters studied and does not have a defined daytime cycle, which leads to the model's incapability of learning its behavior. This variability is largely related to the v-component and not the u-component of the wind, the meridional component (v) has very rapid changes and can pass from

a largely negative value (≈ -1.5) to a high positive value (≈ 1.5) in a few hours, which we hypothesize increases the uncertainty because this is an extra element that the network has to address. For this reason, it is difficult for the LSTM to make accurate predictions related to the wind direction. In contrast, the other variables have a more stable and smoother trend, easier for the model to learn and predict.

In general, the model has good precision to forecast meteorological variables. Nevertheless, it has some flaws that have to be mentioned. The model fails when steeper changes are present in the data, and especially when this kind of changes happen in stations in which they do not happen very often. This feature can be easily seen in the PM$_{2.5}$ or in the wind direction, in a station of low pollutant concentration, when a steeper trend starts the model is not able to represent it. This also happens to the wind direction (due to the v-component): the variable does not have a defined daily trend or tendency in many cases, so the model cannot fully learn the trends and make accurate predictions, a feature that as mentioned in Sect. 3.1 probably could be addressed following Sayeed et al. (2021).

Conclusions

An LSTM model was designed to forecast PM$_{2.5}$ concentrations and meteorological variables. For this purpose, the model was trained with data from 12 AQMNB stations based in Bogotá, a city characterized by its complex topography and meteorology dynamics, located on the Andes Mountains of South America. To evaluate the model, 17 days with an *unhealthy* AQI category and 25 random days were selected. This evaluation was based on statistical parameters recommended for the assessment of air quality modeling results (RMSE, *r*, IOA, MAB, and MGE).

We concluded that the predictions related to PM$_{2.5}$, radiation, temperature, relative humidity, wind speed, and *u* component have excellent and good performance in a city with complex terrain, according to the evaluation parameters and the benchmarks used to compare the ANN-LSTM results with the AQMNB stations data. Furthermore, the model with the lowest accuracy was the one where the *v*-component and the wind direction in degrees are not well represented; this phenomenon may be caused by fast changes in wind dynamics, which is likely due to difficulties of the ANN-LSTM to reproduce the fast changes of the v-component values between positive and negative.

Using the vectorial components of the wind, instead of wind direction in degrees and wind speed scalar values, to simulate wind field is more adequate. This is demonstrated by the better scores of the evaluated statistical parameters compared to the values published in other studies (e.g., Zhao et al. 2019; Guevara-Luna

Fig. 7 Model's performance in terms of the evaluation parameters and benchmark distribution for all 12 stations and the 42 days: correlation (r), RMSE, and IOA for the meteorological variables of radiation, temperature, relative humidity (RH), wind speed, and wind field horizontal components (u and v components of the wind)
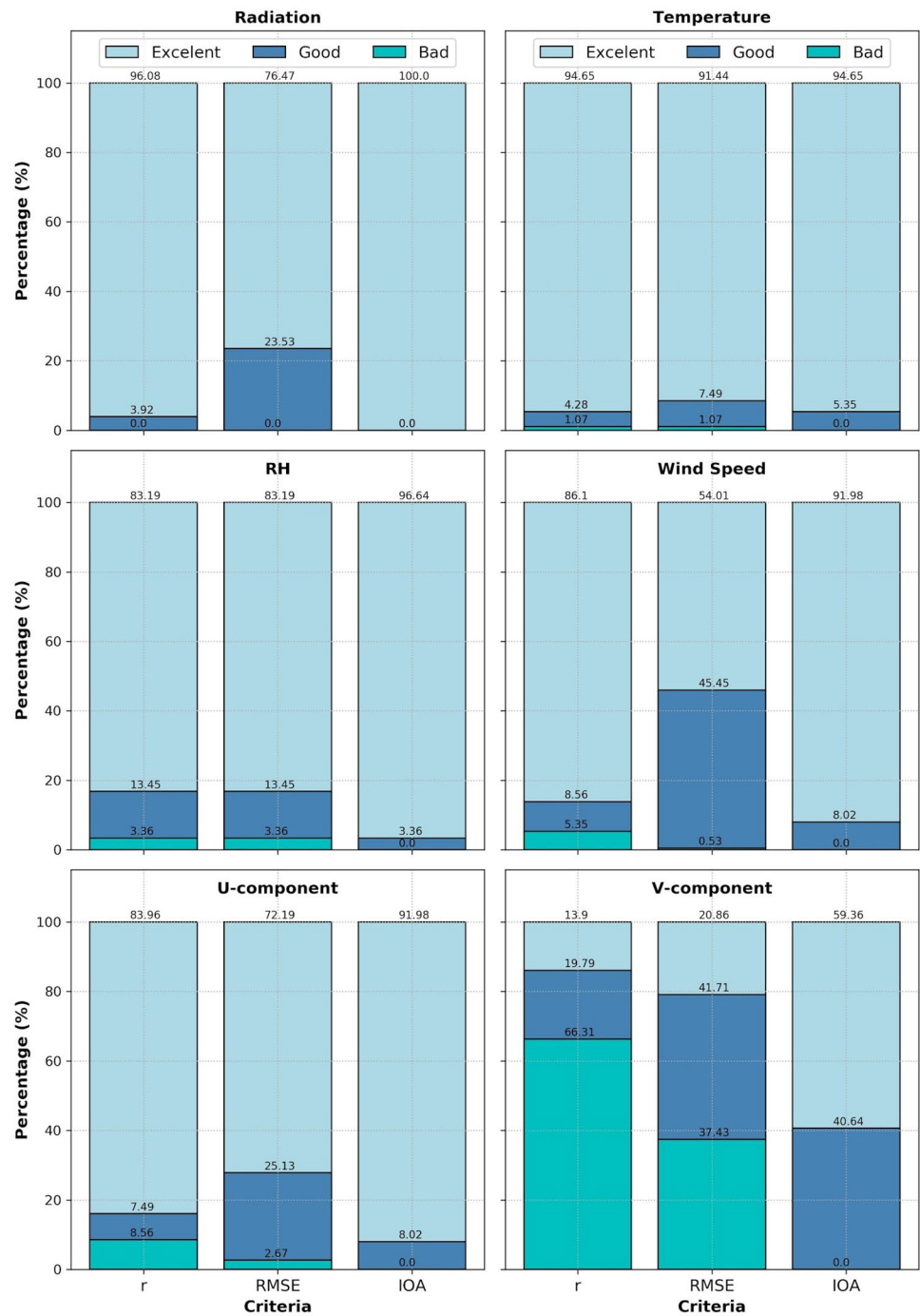
et al. 2020; Mogollón-Sotelo et al. 2020). Nevertheless, the errors are due to the meridional component of the wind (V-component) and can be improved by the addition of a CNN, or by novel approaches such as the merging of the WRF model and neural networks as in Sayeed et al. (2021).

The model can predict the spatio-temporal variation of meteorological variables and PM$_{2.5}$ in Bogotá, delivering good results (i.e., better than other studies, e.g., Kumar et al. 2016; González et al. 2018; Zhao et al. 2019; Casallas et al. 2020; Mogollón-Sotelo et al. 2020-) in terms of

statistical parameters, as demonstrated by the excellent and good results in the benchmarks (Figs. 3, 4, 5). Additionally, a tropical city with complex topography, such as Bogota, was selected to validate the model's capability of performing precise forecasts in places with important and fast variations of wind and thermodynamic fields, and also in its pollution behavior.

Hence, the implemented model has the potential to be used as a tool of risk management, i.e., an early alert system, in other cities with air quality and meteorology monitoring
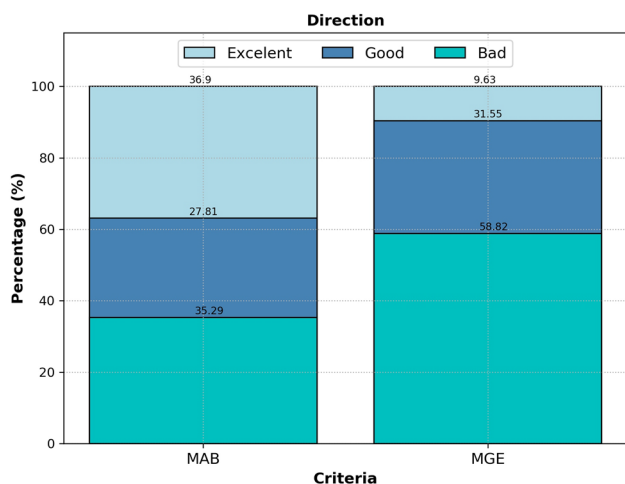
Fig. 8 Model's performance in terms of the evaluation parameters and benchmark distribution for all 12 stations and the 42 days: correlation coefficient (*r*), RMSE, and IOA for the variable of wind direction

Table 3 Interquartile range for the forecasted variables: $PM_{2.5}$ concentration (μg m^{-3}), radiation (W m^{-2}), air temperature (°C), relative humidity (%), wind speed (m s^{-1}) and wind direction (°)

Variable	Interquartile range
$PM_{2.5}$ (μg m^{-3})	10.34
Radiation (W m^{-2})	255.39
Temperature (°C)	1.53
Relative humidity (%)	7.29
Wind speed (m s^{-1})	1.21
Wind direction (°)	93.11

stations that have a record of at least 5 years of hourly data. To use the model as a risk management tool to predict 24 h in advance, the LSTM must be trained every 2 weeks with the latest measured data, which can decrease its errors and ensure the quality of the predictions. This frequent training is required since the LSTM has difficulties forecasting precisely variables with a not clear daily pattern (e.g., wind field), variables in which different temporal components (seasonal, weekly, daily, etc.) may have important variation. This tool based on LSTM has the potential to reduce the negative impact on the health of vulnerable populations and government costs related to public health.

In general, when comparing the LSTM with other machine learning tools and physical models (e.g., Kumar et al. 2016; Casallas et al. 2020; Mogollón-Sotelo et al. 2020), we observed that the implemented model requires less computational time to forecast (it takes the LSTM 1 min to deliver results). This study concludes that the proposed approach (LSTM) has better performance than other

forecasting approaches in terms of precision and trend representation without large computational power needs. This conclusion is based on the results, analysis performed, and evaluation of the model using the statistical parameters of r, MB, RMSE, MGE, MAB, and IOA, for the variables of temperature, radiation, humidity, wind speed, and wind direction. Nevertheless, it is necessary to consider the representation of peaks in variables that do not follow the diurnal trend, due to variability in their behavior (e.g., wind direction). Finally, it is important to highlight that this problem is recurrent with different forecasting techniques; thus, a coupling between machine learning techniques and physical modeling, or other alternative approaches, could help to deal with these inaccuracies.

Availability of data and materials Not applicable.

Declarations

Conflict of interest The authors declare that have no competing interests.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent to publish Not applicable.

References

Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H (2018) State-of-the-art in artificial neural network applications:

a survey. Heliyon 4:e00938. https://doi.org/10.1016/j.heliyon.2018.e00938

Adams D, Seth I, Kirk DS (2013) GNSS observations of deep convective time scales in the Amazon. Geophys Res Lett 40:2818–2823. https://doi.org/10.1002/grl.50573

Baker KR, Foley KM (2011) A nonlinear regression model estimating single source concentrations of primary and secondarily formed PM2.5. Atmos Environ 45:3758–3767. https://doi.org/10.1016/j.atmosenv.2011.03.074

Barten GM, Ganzeveld LN, Visser AJ, Jiménez R, Krol MC (2019) Evaluation of nitrogen oxides sources and sinks and ozone production in Colombia and surrounding areas. Atmos Chem Phys 47:777–780. https://doi.org/10.5194/acp-2019-781

Bengio Y, Simard P, Frasconi P (1994) Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Netw 5:157–166. https://doi.org/10.1109/72.279181

Boylan JW, Russell AG (2006) PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models. Atmos Environ 40:4946–4959. https://doi.org/10.1016/j.atmosenv.2005.09.087

Casallas A, Celis N, Ferro C et al (2020) Validation of PM10 and PM2.5 early alert in Bogotá, Colombia, through the modeling software WRF-CHEM. Environ Sci Pollut Res 27:35930–35940. https://doi.org/10.1007/s11356-019-06997-9

Casallas A, Hernandez-Deckers D, Mora-Paez H (2021) Understanding convective storms in a tropical, high-altitude location with in-situ meteorological observations and GPS-derived water vapor. Atmósfera Early Release. https://doi.org/10.20937/ATM.53051

d'Avila Garcez AS, Broda K, Gabbay DM (2001) Symbolic knowledge extraction from trained neural networks: a sound approach. Artif Intell 125:155–207. https://doi.org/10.1016/S0004-3702(00)00077-1

Emery CA, Tai E, Yarwood G (2001) Enhanced meteorological modeling and performance evaluation for two texas ozone episodes. Final report submitted to Texas Natural Resources Conservation Commission, Prepared by ENVIRON, International Corp., Novato

EPA (2014) AQI—Air Quality Index. A guide to air quality and your health. EPA-456/F-14-002. https://www3.epa.gov/airnow/aqi_brochure_02_14.pdf. Accessed 07 Jan 2020

Erdil A, Arcaklıoğlu E (2013) The prediction of meteorological variables using artificial neural networks. Neural Comput Appl 22:1677–1683. https://doi.org/10.1007/s00521-012-1210-0

Feng R, Zheng H, Gao H, Zhang A, Huang C, Zhang J, Luo K, Fan J (2019) Recurrent Neural Network and random forest for analysis and accurate forecast of atmospheric pollutants: a case study in Hangzhou. China J Clean Prod 231:1005–1015. https://doi.org/10.1016/j.jclepro.2019.05.319

Franceschi F, Cobo M, Figueredo M (2018) Discovering relationships and forecasting PM10 and PM2.5 concentrations in Bogotá, Colombia, using artificial neural networks, principal component analysis, and k-means clustering. Atmos Pollut Res 9:912–922. https://doi.org/10.1016/j.apr.2018.02.006

Freitag BM, Nair US, Niyogi D (2018) Urban modification of convection and rainfall in complex terrain. Geophys Res Lett 45:2507–2515. https://doi.org/10.1002/2017GL076834

Gehling W, Dellinger B (2013) Environmentally persistent free radicals and their lifetimes in PM2.5. Environ Sci Technol 47:8172–8178. https://doi.org/10.1021/es401767m

González CM, Ynoue RY, Vara-Vela A, Rojas NY, Aristizábal BH (2018) High-resolution air quality modeling in a medium-sized city in the tropical Andes: assessment of local and global emissions in understanding ozone and PM10 dynamics. Atmos Pollut Res 9:934–948. https://doi.org/10.1016/j.apr.2018.03.003

Grivas G, Chaloulakou A (2006) Artificial neural network models for prediction of PM10 hourly concentrations, in the Greater Area of Athens. Greece Atmos Environ 40:1216–1229. https://doi.org/10.1016/j.atmosenv.2005.10.036

Guevara-Luna MA, Casallas A, Belalcázar Cerón L et al (2020) Implementation and evaluation of WRF simulation over a city with complex terrain using Alos-Palsar 0.4 s topography. Environ Sci Pollut Res 27:37818–37838. https://doi.org/10.1007/s11356-020-09824-8

Hochreiter S, Schmidhuber J (1997a) Long short-term memory. Neural Comput 9:1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hochreiter S, Schmidhuber J (1997b) Long short-term memory. Neural Comput 9:1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Holton J (2004) An introduction to dynamic meteorology, 4th edn. Elsevier Academic Press, Burlington

Huang CJ, Kuo PH (2018) A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities. Sensors 18:2220. https://doi.org/10.3390/s18072220

Karevan Z, Suykens JA (2020) Transductive LSTM for time-series prediction: an application to weather forecasting. Neural Netw 125:1–9. https://doi.org/10.1016/j.neunet.2019.12.030

Kumar A, Jimenez R, Belalcázar LC, Rojas NY (2016) Application of WRF-Chem model to simulate PM10 concentration over Bogota. Aerosol Air Qual Res 16:1206–1221. https://doi.org/10.4209/aaqr.2015.05.0318

Maqsood I, Khan MR, Abraham A (2014) An ensemble of neural networks for weather forecasting. Neural Comput Appl 13:112–122. https://doi.org/10.1007/s00521-004-0413-4

Martínez NM, Montes LM, Mura I, Franco JF (2018) Machine learning techniques for PM10 levels forecast in Bogotá. In: 2018 ICAI workshops (ICAIW). https://doi.org/10.1109/icaiw.2018.8554995

McNally D (2009) 12km MM5 performance goals. In: The 10th annual ad-hoc meteorological modelers meeting, Boulder, CO, US Environmental Protection Agency. http://www.epa.gov/scram001/adhoc/mcnally2009.pdf. Accessed 5 June 2020

Mogollón-Sotelo C, Casallas A, Vidal S, Celis N, Ferro C, Belalcazar LC (2020) A support vector machine model to forecast ground-level PM2.5 in a highly populated city with a complex terrain. Air Qual Atmos Health 14:399–409. https://doi.org/10.1007/s11869-020-00945-0

Murillo-Escobar J, Sepulveda-Suescun JP, Correa MA, Orrego-Metaute D (2019) Forecasting concentrations of air pollutants using support vector regression improved with particle swarm optimization: case study in Aburrá Valley, Colombia. Urban Clim 29:100473. https://doi.org/10.1016/j.uclim.2019.100473

Ndiaye E, Le T, Fercoq O, Salmon J, Takeuchi I (2019) Safe grid search with optimal complexity. In: Proceedings of the 36th international conference on machine learning, in proceedings of machine learning research, vol 97, pp 4771–4780. http://proceedings.mlr.press/v97/ndiaye19a.html

Önskog J, Freyhult E, Landfors M et al (2011) Classification of microarrays; synergistic effects between normalization, gene selection and machine learning. BMC Bioinform 12:390. https://doi.org/10.1186/1471-2105-12-390

Ostro B, Chestnut L (1998) Assessing the health benefits of reducing particulate matter air pollution in the United States. Environ Res 76:94–106. https://doi.org/10.1006/enrs.1997.3799

Perez P, Gramsch E (2016) Forecasting hourly PM2.5 in Santiago de Chile with emphasis on night episodes. Atmos Environ 124:22–27. https://doi.org/10.1016/j.atmosenv.2015.11.016

Prechelt L (1998) Early stopping - but when? In: Orr GB, Müller KR (eds) Neural networks: tricks of the trade. Lecture notes in computer science, vol 1524. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-49430-8_3

Qing X, Niu Y (2018) Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. Energy 148:461–468. https://doi.org/10.1016/j.energy.2018.01.177

Remm JF, Alexandre F (2002) Knowledge extraction using artificial neural networks: application to radar identification. Signal Process 82:117–120. https://doi.org/10.1016/S0165-1684(01)00142-6

Rojas N (2004) Revisión de las emisiones de material particulado por la combustión de Diesel y Biodiesel. Revista De Ingeniería 20:58–68 (**ISSN 0121-4993**)

Sayeed A, Choi Y, Eslami E et al (2021) A novel CMAQ-CNN hybrid model to forecast hourly surface-ozone concentrations 14 days in advance. Sci Rep 11:10891. https://doi.org/10.1038/s41598-021-90446-6

SDA (2013) Características generales de las estaciones de la Red de Monitoreo de Calidad del Aire de Bogotá y parámetros medidos en cada una de ellas a 2013. Retrieved from http://ambientebogota.gov.co/estaciones-rmcab. Accessed 1 July 2020

Sokhi R et al (2021) A global observational analysis to understand changes in air quality during exceptionally low anthropogenic emission conditions. Environ Int 157:106818. https://doi.org/10.1016/j.envint.2021.106818

Staudemeyer RC, Morris ER (2019) Understanding LSTM—a tutorial into long short-term memory recurrent neural networks. Comput Res Repos. arXiv:1909.09586.

Steinwart I, Christmann A (2008) Support vector machines. Springer, New York

Tong W, Li L, Zhou X et al (2019) Deep learning PM2.5 concentrations with bidirectional LSTM RNN. Air Qual Atmos Health 12:411–423. https://doi.org/10.1007/s11869-018-0647-4

Tuccella P, Curci G, Visconti G, Bessagnet B, Menut L, Park RJ (2012) Modeling of gas and aerosol with WRF/Chem over Europe: evaluation and sensitivity study. J Geophys Res Atmos. https://doi.org/10.1029/2011jd016302

US EPA (United States Environmental Protection Agency) (2000) Meteorological monitoring guidance for regulatory modeling applications. Epa-454/R-99-005, 171. http://www.epa.gov/scram001/guidance/met/mmgrma.pdf. Accessed 26 Apr 2020

UNDRR (2005) Desarrollo de ciudades resilientes: Bogotá. https://www.eird.org/camp-10-15/docs/Bogota.pdf. Accessed 2 July 2020

Vera-Vela A, Andrade MF, Kumar P, Ynoue RY, Munoz AG (2016) Impact of vehicular emissions on the formation of fine particles in the Sao Paulo Metropolitan Area: a numerical study with the WRF-Chem model. Atmos Chem Phys 16:777–797. https://doi.org/10.5194/acp-16-777-2016

Wang X, Sun W (2019) Meteorological parameters and gaseous pollutant concentrations as predictors of daily continuous PM2.5 concentrations using deep neural networks in Beijing–Tianjin–Hebei, China. Atmos Environ 211:128–137. https://doi.org/10.1016/j.atmosenv.2019.05.004

Westerlund J, Urbain JP, Bonilla J (2014) Application of air quality combination forecasting to Bogota. Atmos Environ 89:22–28. https://doi.org/10.1016/j.atmosenv.2014.02.015

Zhao J, Deng F, Cai Y, Chen J (2019) Long short-term memory—fully connected (LSTM-FC) neural network for PM2.5 concentration prediction. Chemosphere 220:486–492. https://doi.org/10.1016/j.chemosphere.2018.12.128

Zhou Y, Chang FJ, Chang LC, Kao IF, Wang YS, Kang CC (2019) Multi-output support vector machine for regional multi-step-ahead PM2.5 forecasting. Sci Total Environ 651:230–240. https://doi.org/10.1016/j.scitotenv.2018.09.111