2019 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)
Mexico City, Mexico. September 11-13, 2019

Evaluation of a Recurrent Neural Network LSTM for the Detection of Exceedances of Particles PM10.

Julio Alberto Ramírez Montañez.
Facultad de Ingenieria.
Universidad Autónoma de Querétaro.
Querétaro, México.
julio.ramirezm27@gmail.com

Marco Antonio Aceves Fernandez.
Facultad de Ingenieria.
Universidad Autónoma de Querétaro.
Querétaro, México
marco.aceves@gmail.com

Saúl Tovar Arriaga
Facultad de Ingenieria.
Universidad Autónoma de Querétaro.
Querétaro, México.
saul.tovar@uaq.mx

Juan Manuel Ramos Arreguin
Facultad de Ingenieria.
Universidad Autónoma de Querétaro
Querétaro, México.
jsistdig@yahoo.com.mx

Giovanni Angelo Salini Calderon
Facultad de Ingenieria.
Universidad Catolica de la Santisima Concepcíon.
Chile
giovanni.angelo@gmail.com

Abstract—**Monitoring air quality is a topic of current interest, since poor quality has a negative impact on health. Air quality is affected by different pollutants, such as particulate matter and gases, produced by the growing industrial development. As a preventive measure, Mexico established different standards in order to control airborne pollution. In this paper, we propose a methodology based upon a recurrent long-term/short-term memory network for the prediction of exceedances of PM10 (particles of less or equal diameter than 10 micrometers) with time intervals of 72, 48 and 24 hours in advance. Obtaining a satisfactory percentage of prediction as a whole a minimum variability in repetitive experimental runs.**

Keywords— Air pollution, neural network (NN), Deep Network.

I. Introduction

The industrial development achieved today has not only brought great benefits and facilities for daily life, but has also caused environmental problems. Air pollution is one of these problems, which seriously affects health, generating various ills [1]. In response, environmental monitoring facilities have been developed, which record the various concentrations of pollutants in a defined time interval. Like several variables that affect these concentrations such as: temperature, relative humidity, wind direction, and the geographical area where the measurement is recorded [2]. The interaction of these variables makes the behavior of the particulate matter highly random, increasing the complexity of modeling their behavior. The particulate material being an air pollutant analyzed in this article. As they are values recorded at defined time intervals, recurrent neural networks have demonstrated their efficiency in modeling the behavior of data recorded with the same capture frequency [3].

First of all, the Particulate material includes all suspended particles, whether its origin is natural or artificial. Therefore, the mechanisms to which these particles are dispersed through the atmosphere are different in comparison to most of the harmful gasses in the atmosphere. Some of these particles are of such a small size that they are able to enter the bloodstream by breathing them, causing serious health problems [1,2].

Some of these particles include: PM10, PM2.5 and ultra-fine. Where PM means "Particulate Matter" and the number then determines the maximum size in microns (μm) as mentioned by Concepción-Jiménez [1]. From the above mentioned classification PM10 particles were selected, as the case study, which are mainly composed of substrates, nitrates, mineral powders, and pollen [4]. In the case of the predictive analysis of the neural network to be used, the focus was on modelling the behaviour of PM10 particles as well as on the characteristics that precede their exceedance.

They are called exceedance when the PM10 particle value exceeds the values defined by some standard. For example, the Mexican standard NOM-025-SSA1-2014 establishes a daily average of 75 $\mu g/m^3$ and 40 $\mu g/m^3$ annually. By definition of the standard, it is necessary to know the registered values of pollutants during a day to determine if the average is below or above the value of the norm, in the same way to know the annual average value.

The main reason for the study of the exceedances is that it has been shown that prolonged exposure to high concentrations of particles in the environment (PM10) is related to increased respiratory and cardiovascular problems. Similarly, the World Health Organization (WHO) has proposed standards, even though it is estimated that 91% of the world population lives in places where the established guidelines are not followed with respect to the established levels [5].

The neural network artificial are inspired by the very neurons that make up a brain, they were developed to simulate the ability to learn, using simple elements. These elements can be classified into input, processing and output elements; all these can be interconnected or not [3,6]. In Figure 1, a schematic of a basic neural network is shown.

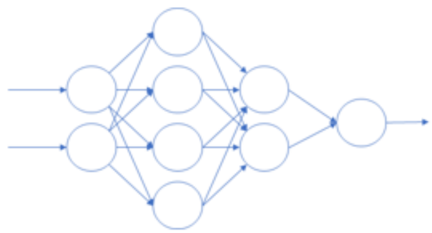

Fig. 1 Example of a four-layer neural network (figure adapted from [3]).

Warren McCulloch and Walter Pitts pioneers in the development of a computer model for an artificial neuron, the basic element of a neuron network, called threshold logic [7]. Over the years, different network topologies have been proposed, each with advantages to solve different problems, whether they are classification, detection, characterization, etc.

In 1997, Hochreiter and Schmidhuder proposed the LSTM model (Long Short-Term Memory), unlike a simple recurrent neural network that has a long-term memory in the form of weights, which are modified during the training of the network and short-term memory defined in activation functions between the communication of the nodes of neurons [8]. The LSTM model introduces a block of internal memory, composed of simple blocks connected in a specific way (Figure 2), each of them are described as follows [9]:

- Input node: represented by "a1" is a node whose activation is usually a sigmoid, which weights the input values.
- Input gate: "a2" is a characteristic of the LSTM. Take the current points in conjunction with the data from a previous step. Managing a logical flag if it is zero cuts the data flow between the nodes and if one is the data pass through it. Acting as a block memory control.
- Internal state: it is the most important part, referenced with an "S", this node has a recurring connection, this connection is usually called a carousel of constant errors. This block prevents the error from increasing.
- Oblivion gate: "a4" introduced by Ger in 2000, provides a method by which the network can adjust to the content of the internal state. Especially for the use of time series.
- Output gate: "output" is the value produced by the value of the internal state multiplied by the value of the output gate "a3", to ensure this output the activation function of the first state is a function tanh.

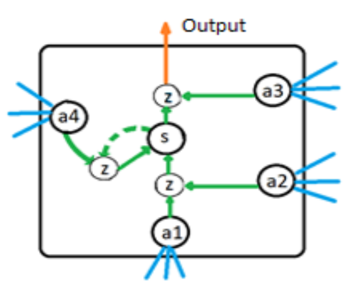

Fig. 2 Basic block of an LSTM (figure adapted form [9]).

Currently, there are few works on the prediction of exceedances of PM10 particles, from the last developed works we found the one made by Park [3], which consisted of the prediction of exceedances based on the records of the previous days, comparing a linear regression, with a LSTM network, using data recorded in Seoul, Korea, between January 2015 and March 2016.

The LSTM network was evaluated with different hidden neuron values in conjunction with different lot size values. The best result obtained was 3.54 of MSE and 1.86 of RMSE, in the modeling of the behavior of PM10 particles, with these results the behavior predicted by the network allowed the prediction of exceedances in a previously defined day [3].

Another work of 2002, the author Brunelli generated a mathematical model for the prediction of exceedances of SO2, being able to predict them three hours in advance [6]. The main disadvantage of the previous work is that the SO2 particles are easily modelled by means of the behavior of the temperature in the environment.

II. Material and methods.

A. Materials.

Mexico has a constant monitoring network in Mexico City (CDMX) called the Atmospheric Monitoring Network (RAMA). This network has 24 stations distributed in different municipalities. Each station registers in each hour the levels of different atmospheric pollutants such as the level of ozone, nitrogen dioxide, nitrogen oxides, nitrogen monoxides, sulphur dioxide, carbon monoxides, PM10, PM2.5, among others [10].

This database is available on RAMA [10]. In this work used the data from the Merced monitoring station (MER), from the year 2000 to 2018 for training purposes. In Figure 3, it is observed its location being a central point.

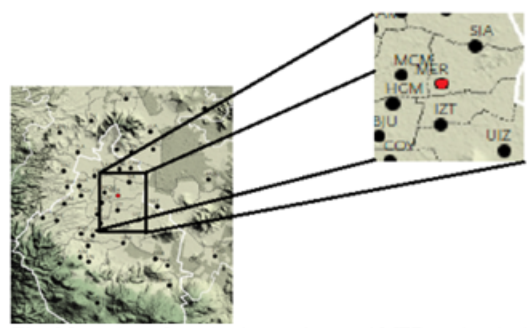

Fig. 3 Location of monitoring stations and MER station (red color) (figure adapted from 11).

B. Model construction.

Before defining the structure of the network, it is necessary to scale the training data in a range of -1 to 1, with the purpose of decreasing the non-linearity of the initial data, facilitating the training of the LSTM neural network.

- Modeling neural network.

This network has 50 neurons in the input layer followed by a hidden layer of 256 neurons and ends with a complete connection layer to which it produces the output vector [12].

- ❖ Layer 1 consists of 50 LSTM neurons, which will take the first 50 data and expand them to feed the second layer, generating a record of them.
- ❖ Layer 2 consists of 256 LSTM neurons.

- ❖ Layer 3 is a simple neuron, which based on the previous recorded data will generate a new value, successively.
  - Predictive neural network.

For this the number of incoming neurons is determined by the length of the input vector determined by the number of values prior to evaluating, since 24, 48 and 72 hours, each one represents the previous hours registered to an exceedance and days without exceedances. In its second layer, it is an output layer of a neuron.

- ❖ Layer 1 each neuron receives a value from the input data vector, which generates an output response.
- ❖ Layer 2 receives the results of the first network and generates a classification, given that initially the days with and without exceedances are known, the network uses a continuous regression to adjust its weights and obtain the expected result.
  - Complete model.

Figure 4 shows the connection between the two neural networks; the first three layers are from the modeling network of the behavior of the PM10 and the next two layers are from the classification network. Once the two networks previously included are present, the first network supplies a previously established length vector (between 72, 48 and 24) and the second network provides a flag if the day after the length of the vector will be an exceedance or not.

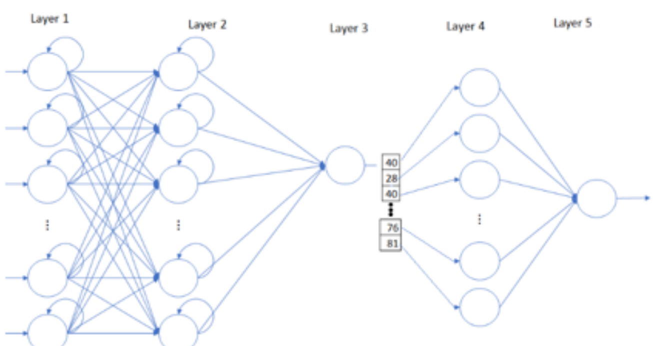

Fig. 4 Complete model.

C. Evaluation of the model.

To evaluate the correct modeling generated by the LSTM network, the Mean Square Error (MSE) and the Root Mean Square Error (RMSE) were implemented, the MSE measures the average of the squared errors, in other words, it measures the difference of the real data and the estimates, its equation is represented below:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(YM_i - YR_i)^2 \quad (1)$$

Where: YM is the vector of values modeled by the LSTM network, YR represents the vector of the actual comparison values, and n represents the length of the analyzed vector, both the YM and YR vector +must be of the same length

Meanwhile, the RMSE represents the standard deviation between actual and predicted values, and it is simply to apply a square root to the MSE, generating the following equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(YM_i - YR_i)^2} \quad (2)$$

D. Methodology.

The following diagram (Figure 5) describes the methodology used during the development of this work.

The study data were acquired from the RAMA database of Mexico City, from January 2000 to December 2018. Due to the industrial and residential development of the city, several monitoring stations were incompatible or removed from the network during the course of the years, the first step is to detect the constant stations from 2000 to date, thus reducing the possible stations for the study. Subsequently, these records were evaluated, in order to detect that station with the highest number of valid records with which to train the neural network.

In the sections where the data is not valid, it is proposed to replace it with values generated by an average between the data before and after said section, in order to continue in a certain way with the behavior of PM10 particles. Before supplying the values to the network, it was necessary to generate a single data vector, in order to have a continuous vector from 2000 to 2017.

The LSTM network uses an 80-20 ratio for training and validation, 80 represents the percentage of data used for training while the remaining 20 is used for the validation of the model, the results will be evaluated through the MSE and the RMSE.

In a parallel cycle the original data will be grouped into sections of 24 data, this will allow to know the average recorded individually for the days from 2000 to 2015. Identifying the days with exceedances and also keeping a vector of the conditions prior to the registration of said leave. These records will be used to train the second network.

Having the modeling of PM10 particles provided by the first network and the second being able to anticipate the day of leave, the years 2017 and 2018 were used to evaluate both networks and observe their ability to predict together.

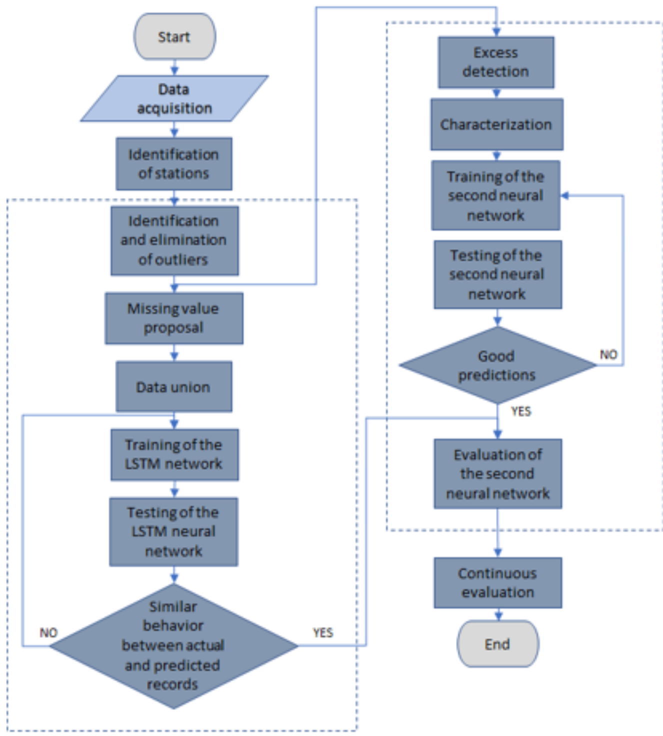

Fig. 5 Methodology used.

III. Results and Discussions.

The recurrent network LSTM shows a mean root mean square error (RMSE) of 18.0342, both are relatively good results, in Figure 6 we can appreciate the behavior of the data recorded by the station, blue and orange color shows the behavior predicted by the network, it may not reach the highest peaks but you can perceive a similarity in the behavior of both graphs.

By increasing the scale of the graph, it is possible to observe the similarities in the behavior of both registers, in Figure 7 the behavior can be observed in the first 500 hours of the year.

In the areas where the sensor had measurement problems, registering outliers, the predicted values are able to complete a behavior similar to the one that has no valid data.

In Figure 8 the exceedances registered by the monitoring station are observed in blue, while in orange the ones registered by the data supplied by the LSTM network and evaluated by the classification network are shown. The green line shows the maximum value allowed by the Mexican norm.

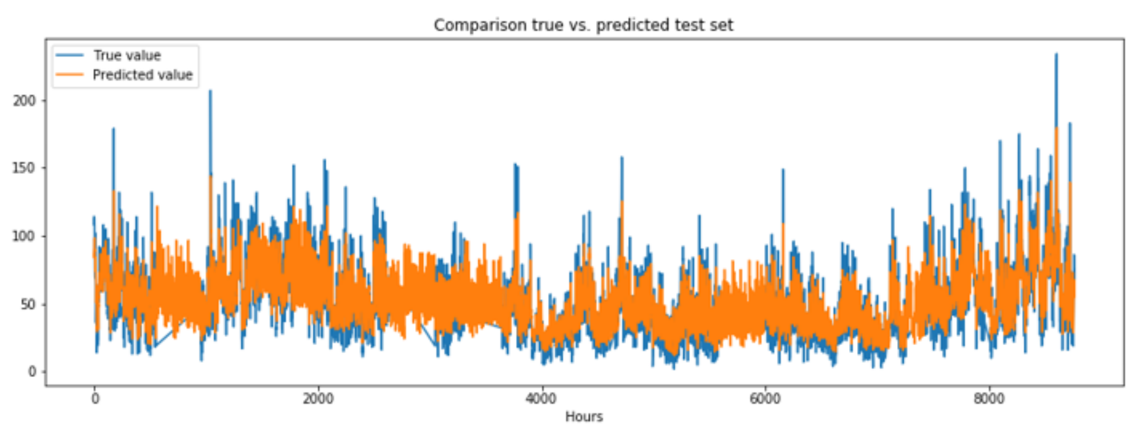

Fig. 6 Comparison between true value and predict value.

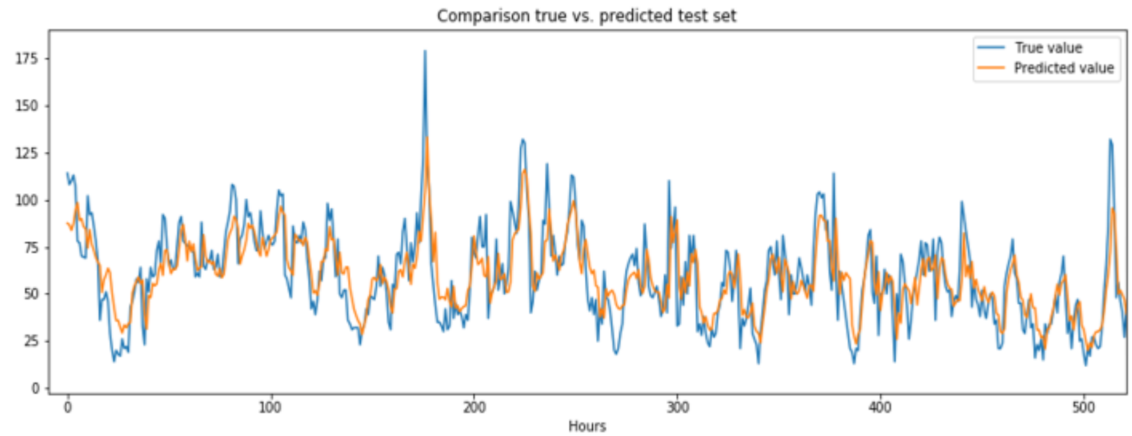

Fig. 7 Increased scale.

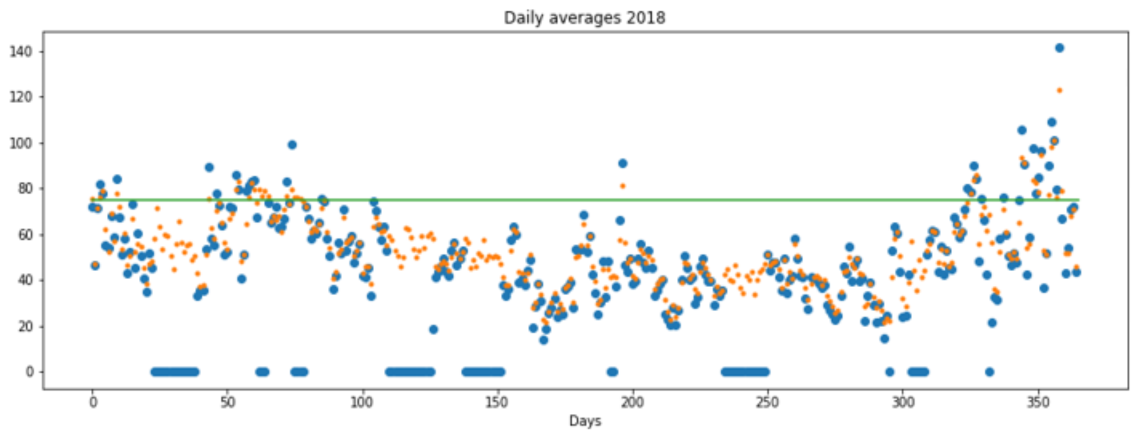

Fig. 8 Comparison of actual and predicted surpluses by the neural network (self-image).

The results of the prediction vary depending on the anticipation time with which work is carried out, the shorter the time the more stable the predictions are when evaluating repetitively. However, 72 hours is perceived in occasions considerably lower, obtaining 9% prediction for a prediction. The reason for this, needs further investigation. In the three time intervals have a steady result of around 89% prediction, which shows consistency for most test runs. In Figure 9, it is shown the variations obtained by evaluating the network 50 times.

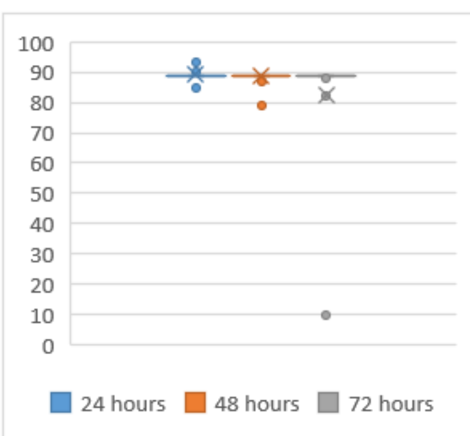

Fig. 9 Variations in the prediction percentage.

IV. Conclusions.

With the present work, the great versatility of an LSTM network and a simple neural network for the prediction of exceedances in an interval of 48 and 24 hours previous was demonstrated. In the case of 72 hours in advance, it will be crucial to investigate the reasons for some predictions lower than expected and improve upon the modeling of the behavior of the PM10 particles in order to reduce or eliminate the low percentages of prediction obtained regularly.

Alternatives of solution can be several, initially it is possible to revise the proposal of the missing data not registered by the stations, in this work an average was made between the behavior before and after the registration of outliers. Another solution could be to generate an average by comparing the behavior at the precise moment in which the sensor recorded the -99 with the value recorded previous or later years.

The next step would be to review the training of the LSTM network, with numerous possible combinations in the parameters that characterize the network architecture.

References

[1] Concepción Jiménez, M. (2015). Definition and validation of a methodology to correlate concentration of atmospheric pollutants and hospital admissions. (Master's Thesis). University of Seville, Seville.

[2] Stadlober, E., Hörmann, S., & Pfeiler, B. (2008). Quality and performance of a PM10 daily forecasting model. Atmospheric Environment, 42 (6), 1098-1109.

[3] Park, J.-h., Lee, K.-h., Kim, K.-j., y Son, U.-h. (2017). PM10 Density Forecast Model Using Long Short Term Memory, 576-581.

[4] Air quality and health. Obtained from the World Health Organization. Retrieved from www.who.int/en/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health.

[5] OMS. (2018a). Calidad del aire y salud. urlhttp://www.who.int/es/news-room/factsheets/detail/ambient-(outdoor)-air-quality-and-health. (Accedido 20-11-2018).

[6] Brunelli, U., Piazza, V., Pignato, L., Sorbello, F., y Vitabile, S. (2008). Three hours ahead prevision of SO2pollutant concentration using an Elman neural based forecaster. Building and Environment, 43 (3), 304-314. doi:10.1016/j.buildenv.2006.05.011.

[7] Hassoun, M. H. (1995). Fundamentals of artificial neural networks. MIT press.

[8] Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv: 1506.00019.

[9] Hua, Y., Zhao, Z., Li, R., Chen, X., Liu, Z., & Zhang, H. (2019). Deep Learning with Long Short-Term Memory for Time Series Prediction. *IEEE Communications Magazine*.

[10] Databases - Automatic Air Monitoring Network (RAMA). Obtained from the World Health Organization. Retrieved from www.aire.cdmx.gob.mx/default.php?opc=%27aKBh%27

[11] Monitoring stations. Obtained from www.aire.cdmx.gob.mx/default.php?opc=%27ZaBhnmI=&dc=%27ZA==

[12] Chen, K., Zhou, Y., & Dai, F. (2015, October). A LSTM-based method for stock returns prediction: A case study of China stock market. In *2015 IEEE International Conference on Big Data (Big Data)* (pp. 2823-2824). IEEE.

[13] Lenschow, P., Abraham, H. J., Kutzner, K., Lutz, M., PreuB, J. D., & Reichenbächer, W. (2001). Some ideas about the sources of PM10. Atmospheric Environment, 23-33

[14] Brunekreef, B., & Holgate, S. T. (2002). Air pollution and health. *The lancet*, *360*(9341), 1233-1242.