

# ASSESSING RECURRENT AND CONVOLUTIONAL NEURAL NETWORKS FOR TROPOSPHERIC OZONE FORECASTING IN THE REGION OF VITÓRIA, BRAZIL

AMILTON S. R. JUNIOR, ERICK GIOVANI SPERANDIO NASCIMENTO  
& DAVIDSON MARTINS MOREIRA  
SENAI CIMATEC, Brazil

## ABSTRACT

The objective of this work is to verify the viability of using recurrent and convolutional neural networks models in the task of predicting ozone levels in the troposphere every hour for the next 24 hours. For this, the recurrent neural networks long short-term memory (LSTM), gated recurrent unit (GRU), and convolutional neural networks (CNN) models were trained using data from an air quality monitoring system collected in the metropolitan region of Vitória, Espírito Santo, Brazil, which contains several meteorological and air quality parameters. For validation, we applied the models over unseen data and compared different neural networks architectures using results obtained according to variations in hyperparameters such as the lookback and the number of layers, in addition to comparing the results obtained from the neural networks models applied with those obtained by a deep multi-layer perceptron, available in the literature. We show that the use of recurrent neural networks is a viable alternative for the task of predicting ozone levels and the LSTM model had the best results amongst the applied models.

*Keywords:* ozone forecasting, air quality, neural networks, deep learning, recurrent neural networks, convolutional neural network.

## 1 INTRODUCTION

Ozone ( $O_3$ ), formed from photochemical reactions together with other pollutants such as nitrogen oxides ( $NO_x$ ) and volatile organic compounds (VOC), is the most important photochemical oxidant in the troposphere and, according to reports based on studies and released by the World Health Organization (WHO), tropospheric ozone is recognized as a threat to human health [1].

This is an oxidizer that can make intercontinental trips, as it is a pollutant that can remain in the free troposphere for weeks, but when in urban regions can rapidly be consumed or formed, having a short lifetime due to its precursors that are emitted by the large number of vehicles and industries. It is also related to respiratory problems: it is estimated that around 5–10% of all deaths related to air pollution come from exposure to ozone [2]. In addition to the human threat, it is possible to observe environmental impacts in the loss of biomass in forests that is caused by the negative influence that ozone can cause on them [3].

Forecasting ozone levels in the troposphere for a specified number of hours ahead can be very helpful in assisting research and precautionary measures involving air quality and human health, in operational forecasting systems. The 2005 WHO air quality guidelines sets the recommended value for ozone concentration for a daily maximum 8-hour average [4]. It is important to develop a predictive model that exceeds this limit to allow immediate preventive responses to health threats. Thus, the objective of the present work is to study and develop different types of deep neural networks aiming to build a model capable of predicting the levels of tropospheric ozone hourly up to the next 24 hours, with good performance. For this, the recurrent neural networks long short-term memory (LSTM), gated recurrent unit (GRU) models, and the convolutional neural network (CNN) model were studied and developed for the task of prediction. A dataset, which contains various meteorological and



air quality data of the region of Vitória, state of Espírito Santo, Brazil, was used for training and validating the aforementioned models. The metrics mean squared error (MSE), mean absolute error (MAE), Pearson  $r$  and  $R^2$  were used to evaluate the results of the models. The results obtained by the models were then compared with a multi-layer perceptron model presented in the literature.

This present work is divided into five main sections. Section 2 addresses related works and explores a variety of works whose context is within the use of techniques for predicting ozone levels. Section 3 contains basic concepts of how the LSTM, GRU and CNN neural network models are implemented in general. Section 4 shows details of the structure of the data used in the region of study and the topologies of the developed neural network models. Section 5 presents the results obtained from the metrics of the trained models and contains the analysis of these results together with their comparison with the results of an MLP model found in the literature. Finally, Section 6 contains final considerations and discussions related to future work.

## 2 RELATED WORKS

A variety of techniques for studying ozone levels are widely disseminated and can be found in several works in the literature [5]–[10].

Within the field of artificial neural networks, it is important to mention the work of Sousa et al. [5], in which the authors used a supervised learning approach to train models of multiple linear regressions and multi-layer perceptron (MLP) for this task of forecasting ozone levels in the troposphere every hour for the next 24 hours. The models were trained using principal components as input, and used a dataset with data collected from an urban site located in Porto, Portugal. As a result, the authors showed that the artificial neural network model presented better accuracy for the prediction of ozone levels compared to linear models because of the non-linearity of the addressed problem. In addition, the authors showed that the use of principal components as input of the neural network generated better results than the use of the original data only, as it reduced the complexity of the model.

The work of Gomez-Sanchis et al. [6] addresses the use of artificial neural networks, using MLP specifically, for the estimation of ozone levels in the troposphere from input data related to meteorological variables and vehicle emissions obtained from the south of the city of Valencia, Spain. The authors' study focuses on finding the relevance among climatological variables and atmospheric contaminants for the prediction of ozone levels. The authors showed the high complexity of the variables that determine ozone levels in the troposphere, stating that they are non-linear and time-dependent, but emphasize that few variables were used.

The use of an MLP model to predict ozone levels in the troposphere is also found in the work of Alves et al. [7], where the authors developed, trained and validated a predictive model using data from an Air Monitoring System located in the city of Vitória, Brazil. The authors used several meteorological and air pollution variables during the development of the work, aiming to forecast the average tropospheric ozone levels for each hour, up to 24 hours.

There are also works that use recurrent neural networks for the forecasting of ozone levels, such as the work of Biancofiore et al. [8], in which the authors used a recurrent three-layer architecture of an Elman artificial neural network to predict ozone levels every hour for the next 48 hours. This architecture presented by the authors can be summarized as a recurrent multi-layer perceptron, which is an extension of a feedforward MLP [9]. The authors used a dataset from the city of Pescara, Italy, containing several variables, some of which are:  $O_3$ ,  $NO_2$ ,  $NO$ ,  $NO_x$  levels, relative humidity and wind speed and direction. As a result, the authors



showed that the recurrent neural network had better results compared to feed-forward neural networks.

Within the context of recurrent neural networks, the work of Pak et al. [10] addresses the use of a hybrid model of long short-term memory and convolutional neural network for the task of ozone levels forecasting in the city of Beijing, China. According to the authors, the CNN model is able to extract efficiently the characteristics of large amount of data related to air quality and weather, while the LSTM is able to reflect the long-term historical process contained in the data.

### 3 RECURRENT AND CONVOLUTIONAL NEURAL NETWORKS

The artificial neural networks used in this work were the CNN, and the recurrent networks called GRU and LSTM. An overview of these types of neural networks will be shown in this section.

#### 3.1 Long short-term memory (LSTM)

Fig. 1 shows the structure of an LSTM cell. All lines carry data that can go through pointwise operations, neural network layers, concatenations and replications.

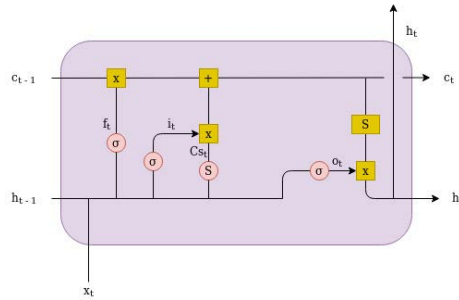


Figure 1: Structure of an LSTM cell. Arrows represent data flow, squares represent pointwise operations, and circles correspond to the activation functions.

LSTM is a special type of recurrent neural network that was created as an approach to the problem of forgetting long-term dependencies common in recurrent neural networks. LSTM was first introduced by Hochreiter and Schmidhuber [11] and, since then, many variants of this neural network have been created.

The most important part of an LSTM is the cell state, which is an internal selective memory of the past, represented by the horizontal line that starts at  $c_{t-1}$  and ends at  $c_t$ . The hidden state, represented by  $h$ , is the output of the LSTM cell. Eqns (1)–(6) represent the functions seen in Fig. 1, given by

$$f_t = \sigma(W_f \cdot [h_{(t-1)}, x_t] + b_f), \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{(t-1)}, x_t] + b_i), \quad (2)$$

$$C_{st} = S(W_C \cdot [h_{(t-1)}, x_t] + b_C), \quad (3)$$

$$C_t = f_t * C_{(t-1)} + i_t * C_s, \quad (4)$$

$$o_t = \sigma(W_o \cdot [h_{(t-1)}, x_t] + b_o), \quad (5)$$

$$h_t = o_t * S(C_t), \quad (6)$$

where  $f_t$  represents the forget gate,  $i_t$  represents the input gate,  $Cs_t$  represents the candidate for the cell state at a given timestep  $t$ ,  $C_t$  represents the cell state at a given timestep  $t$ ,  $o_t$  represents the output gate at a given timestep  $t$ ,  $\sigma$  represents the sigmoid operation,  $S$  represents a hyperbolic tangent operation,  $W_x$  represents the weight for the respective gate  $x$  neurons,  $h_t$  represents the output of the LSTM block at a given timestep  $t$ .  $x_t$  represents the input at the current timestep and  $b_x$  represents the biases for the respective gates  $x$ .

### 3.2 Gated recurrent unit (GRU)

The GRU, as seen in Fig. 2, is a variation of the LSTM that was introduced by Cho et al. [12].

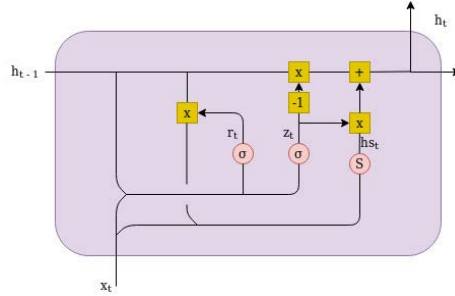


Figure 2: Structure of a GRU cell. Arrows represent data flow, squares represent pointwise operations, and circles represent activation functions.

GRU merges the input and forget gates, in addition to also merging the cell and the hidden states, which generates a simpler model than the LSTM. GRU is able to perform as well as LSTM in certain tasks, especially in small datasets, but LSTM is still more powerful in overall [13]. The GRU's functions are described in eqns (7)–(10), as follows

$$z_t = \sigma(W_z \cdot [h_{(t-1)}, x_t]), \quad (7)$$

$$r_t = \sigma(W_r \cdot [h_{(t-1)}, x_t]), \quad (8)$$

$$h_{st} = S(W \cdot [r_t * h_{(t-1)}, x_t]), \quad (9)$$

$$h_t = (1 - z_t) * h_{(t-1)} + z_t * h_{st}, \quad (10)$$

where  $z_t$  represents the update gate,  $r_t$  represents the reset gate,  $h_{st}$  represents the current memory content and  $h_t$  represents the final memory content.

### 3.3 Convolutional neural networks (CNN)

Fig. 3 shows a graphical representation of a typical CNN architecture.

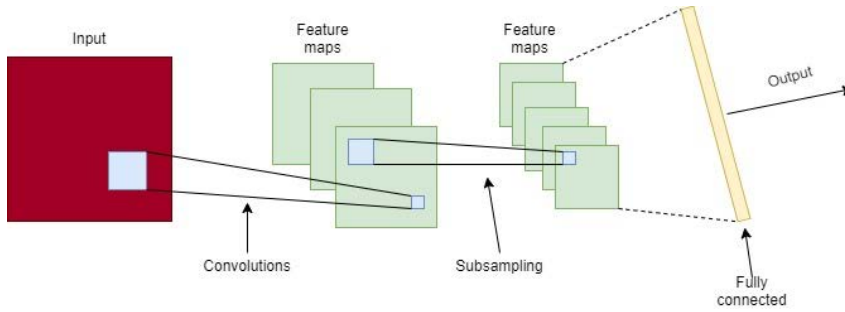


Figure 3: Typical CNN architecture.

A CNN is a type of artificial neural network that applies filters to the input data and has great learning power for spatial and even temporal relationships within the input data. CNN is able to reduce data size through its convolution operations conducted by the kernel/filter in the convolution layer, which extracts high-level features from the data. The kernel convolution is a process where the input is transformed based on the values of a filter (eqn (11)). Eqns (12) and (13) describe the convolution layer

$$G[m, n] = (f * k)[m, n] = \sum_j \sum_i k[j, i] f[m - j, n - i], \quad (11)$$

$$V^l = K^l \cdot C^{[l-1]} + b^l, \quad (12)$$

$$C^l = a^l(V^{[l]}), \quad (13)$$

where  $G$  represents the feature map,  $f$  represents the input,  $k$  represents the kernel,  $m$  and  $n$  represent the rows and columns of the result matrix,  $j$  and  $i$  represent the elements of the filter kernel,  $l$  represents the layer position,  $V$  represents the intermediate value,  $K$  represents the tensor that contains filters,  $C$  represents the result of the convolution of the input data,  $b$  represents the bias and  $a$  represents the activation function.

There is also the pooling layer, which, in order to reduce the necessary processing power required for the network training, is responsible for performing a reduction of the spatial size of the output of the convolution step. Example of pooling operations are the maximum value (MaxPooling) and the average value (AvgPooling) in a certain fixed-size window. Soon after, there is a layer responsible for flattening the output so that a regular MLP can use the main features extracted from the input by the previous layers to perform the desired task of regression or classification. Because of these characteristics, CNNs are widely used in computer vision to perform tasks such as, for example, detecting objects in images.

#### 4 METHODS AND DATA

The dataset used for training and validating the models developed in this work consists of data from an air monitoring system of the State Institute of Environment of Espírito Santo (Instituto Estadual de Meio Ambiente e Recursos Hídricos do Espírito Santo – IEMA – ES). The data were collected during the period between 2001 and 2015 in the region of Cariacica, which is a city in the metropolitan region of Vitória, Espírito Santo, Brazil (Fig. 4). This data is available for public use.

The dataset consists of hourly averages of air pollutant and meteorological measurements, namely: total suspended particulate matter (PM<sub>10</sub>), sulfur dioxide (SO<sub>2</sub>), nitrogen monoxide (NO), nitrogen dioxide (NO<sub>2</sub>), NO<sub>x</sub>, carbon monoxide (CO), O<sub>3</sub>, air temperature and



Figure 4: Map with locations of air monitoring stations. Numbers represent each station. (Source: <https://iema.es.gov.br/>.)

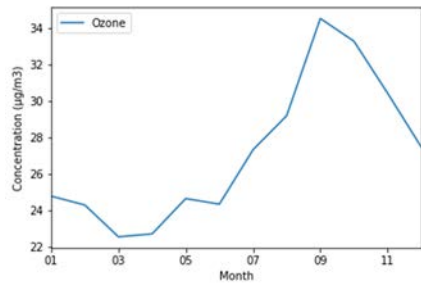


Figure 5: Monthly average of tropospheric ozone concentrations in the period between 2001 and 2005 in the region of Cariacica, Vitória, Brazil.

humidity, scalar wind direction and wind speed. On average, the ozone concentrations starts to increase in June, reaching the highest values in September (Fig. 5). The highest ozone concentrations occurred in 2015 with a value of  $126 \mu\text{g}/\text{m}^3$ .

For the purpose of comparing the results, the data used in this work for training, validation and testing were the same as the work of Alves et al. [7], in which a data treatment in the presented dataset was carried out. The data was scaled to a range between 0 and 1, eliminating missing or invalid data, including all data from the years of 2007, 2009, 2010 and 2013, while maintaining the consistency of the time series. Therefore, this work used 30,492 samples, where 22,987 data points from the years 2001 to 2005 were used for training and validation, and 7,505 data points from the years 2006, 2008, 2011, 2012, 2014 and 2015 were used for testing the models. The authors split these two batches of data so that they both have samples for all periods in different time scales (hours, days, weeks and months). The metrics used for the evaluation of the models performance were the mean squared error (MSE), mean absolute error (MAE), Pearson  $r$ ,  $R^2$  and mean absolute percentage error (MAPE).

The topology of the developed LSTM neural network is shown in Table 1. In it, 4 hidden LSTM layers were created with 94 neurons in the first layer, 55 neurons in the second layer, 36 in the third layer and 48 in the fourth layer. The activation function used was the rectified linear unit (ReLU) for all layers and the optimization function was Adadelta. The loss function used was the MSE.

Table 1: Developed LSTM architecture.

Layer	LSTM topology		
	Layer type	Neurons	Activation function
1st Hidden layer	LSTM	94	ReLU
2nd Hidden layer	LSTM	55	ReLU
3rd Hidden layer	LSTM	36	ReLU
4th Hidden layer	LSTM	48	ReLU
Output	Dense	24	Linear

The topology of the GRU neural network was similar to LSTM and can be seen in Table 2. The only difference between the LSTM and GRU models was the use of GRU layers in the latter.

Table 2: Developed GRU architecture.

Layer	GRU topology		
	Layer type	Neurons	Activation function
1st Hidden layer	GRU	94	ReLU
2nd Hidden layer	GRU	55	ReLU
3rd Hidden layer	GRU	36	ReLU
4th Hidden layer	GRU	48	ReLU
Output	Dense	24	Linear

The topology of the CNN can be seen in Table 3. For this neural network, a one-dimensional convolutional layer was created, followed by a MaxPooling and a flatten layer. The convolutional layer has 128 filters and its activation function was ReLU. The optimization function used in the model was also Adadelta and the loss function used was MSE. The output layer of all implemented models has 24 neurons, which represent all 24 hours of the forecasting, with a linear activation function. Data referring to five timesteps were used as input for the task of predicting the 24 future timesteps.

Table 3: Developed CNN architecture.

Layer	CNN topology		
	Layer type	Filters/neurons	Activation function
1st Hidden layer	1D convolution	128	ReLU
2nd Hidden layer	1D MaxPooling	N/A	N/A
3rd Hidden layer	Flatten	N/A	N/A
Output	Dense	24	Linear

## 5 RESULTS AND DISCUSSION

As results, Table 4 presents the evaluation of the metrics obtained for all models developed in this work and the MLP presented in the work of [7]. These results are related to the forecasting of ozone levels every hour for the 24 hours ahead. It is possible to note that the models with the best results are the LSTM and the MLP, both being comparable, although the LSTM is slightly worse than the MLP when considering all 24 hours ahead.



Table 4: Models metrics results comparison.

MODEL	MSE	MAE	r	R <sup>2</sup>	MAPE
LSTM	102.3	7.69	0.77	0.59	72.8
GRU	149.7	9.67	0.61	0.38	105.6
CNN	156.1	9.79	0.61	0.38	105.6
MLP	101.8	7.68	0.77	0.59	70.5

However, as we can see in Tables 5 and 6, it is notable that the LSTM presents better results of MSE and MAE when it tries to predict more than 9 hours ahead, with the exception of the 24th hour, where the MLP presents better results. This shows the potential of LSTM to learn long-term dependencies in the forecasts.

Table 5: Multilayer perceptron performance metrics. Results taken from [6].

Timesteps	MSE	MAE	r	R <sup>2</sup>	MAPE
T + 1	42.90	4.94	0.91	0.83	41.14
T + 2	65.37	5.93	0.86	0.75	51.10
T + 3	77.46	6.61	0.83	0.69	57.65
T + 6	99.42	7.56	0.77	0.60	67.39
T + 9	110.5	8.01	0.75	0.56	72.44
T + 12	114.1	8.19	0.74	0.55	75.13
T + 15	113.5	8.23	0.74	0.55	76.20
T + 18	111.7	8.16	0.74	0.56	75.81
T + 21	106.7	7.97	0.76	0.57	75.84
T + 24	101.5	7.77	0.77	0.59	76.73

Table 6: Long short-term memory performance metrics.

Timesteps	MSE	MAE	r	R <sup>2</sup>	MAPE
T + 1	60.58	5.68	0.87	0.75	45.63
T + 2	78.38	6.62	0.82	0.68	58.07
T + 3	87.09	6.99	0.80	0.64	60.12
T + 6	101.8	7.62	0.77	0.59	67.28
T + 9	110.7	8.05	0.74	0.56	75.91
T + 12	112.2	8.09	0.74	0.55	76.67
T + 15	109.1	8.01	0.75	0.56	77.0
T + 18	105.4	7.89	0.75	0.57	79.02
T + 21	103.4	7.78	0.76	0.58	78.18
T + 24	108.1	8.04	0.75	0.56	83.2

The individual results of GRU and CNN are presented in Tables 7 and 8.

Fig. 6 shows a graphical comparison of the MSE between the models for each hour predicted within the 24-hour period.

Fig. 7 represents the same comparison, but considering the R<sup>2</sup>. All metrics presented in the tables were calculated according to the original scale of the data.

Figs 8 and 9 show comparisons between the values observed and predicted by the LSTM for predictions of the next hour and for the same hour on the next day in certain periods.



Table 7: Gated recurrent unit performance metrics.

Timesteps	MSE	MAE	r	R <sup>2</sup>	MAPE
T + 1	156.9	9.93	0.59	0.35	111.6
T + 2	150.4	9.65	0.62	0.38	106.4
T + 3	155.2	9.82	0.59	0.35	103.8
T + 6	148.7	9.64	0.61	0.38	108.2
T + 9	152.4	9.73	0.60	0.36	105.5
T + 12	153.7	9.87	0.60	0.36	112.2
T + 15	151.0	9.70	0.60	0.37	105.4
T + 18	153.6	9.84	0.61	0.37	107.8
T + 21	142.0	9.43	0.65	0.42	103.0
T + 24	141.9	9.44	0.64	0.41	102.7

Table 8: Convolutional neural network performance metrics.

Timesteps	MSE	MAE	r	R <sup>2</sup>	MAPE
T + 1	152.7	9.66	0.61	0.37	93.9
T + 2	167.8	10.12	0.57	0.32	107.7
T + 3	148.6	9.49	0.62	0.39	93.5
T + 6	151.5	9.66	0.62	0.39	94.1
T + 9	167.3	10.24	0.57	0.33	110.1
T + 12	173.8	10.45	0.57	0.32	109.5
T + 15	130.3	8.88	0.68	0.46	92.4
T + 18	150.7	9.56	0.64	0.41	88.9
T + 21	170.3	10.29	0.59	0.35	107.3
T + 24	172.2	10.27	0.56	0.31	112.6

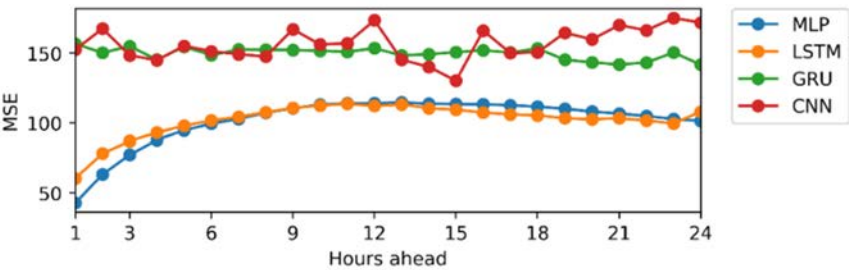


Figure 6: Graphical comparison of the MSE between the models for each hour predicted within the 24-hour period.

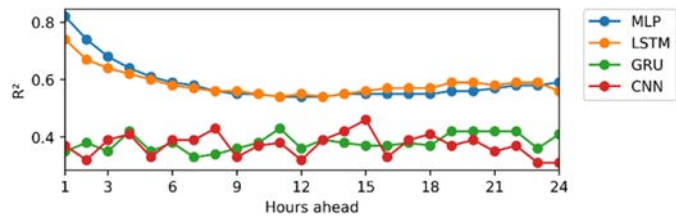


Figure 7: Graphical comparison of the  $R^2$  between the models for each hour predicted within the 24-hour period.

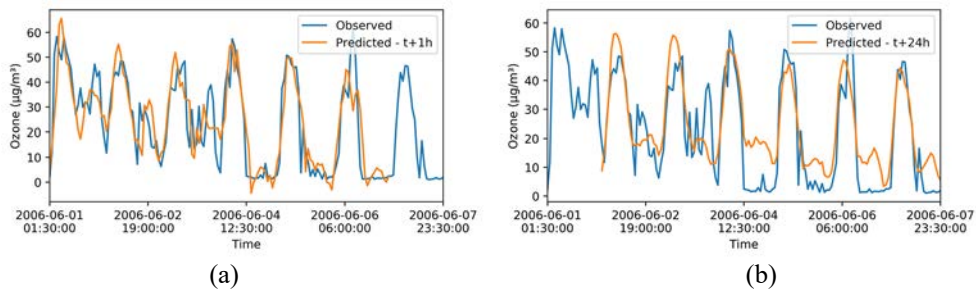


Figure 8: Long short-term memory predictions for a 7-day period in June 2006. (a) Predictions 1 hour ahead; and (b) Predictions 24 hours ahead.

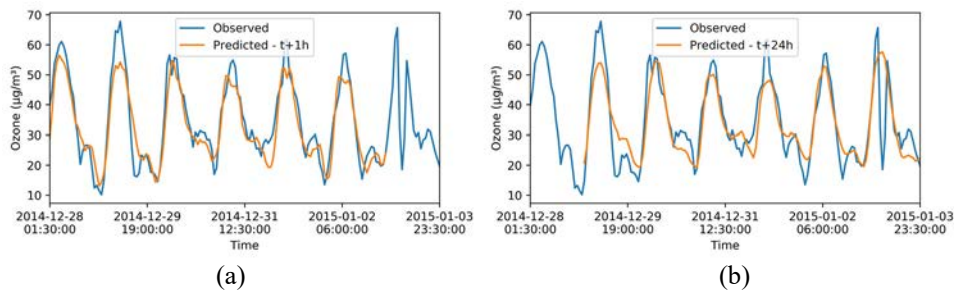


Figure 9: Long short term memory predictions for a 7-day period between December 2014 and January 2015. (a) Predictions 1 hour ahead; and (b) Predictions 24 hours ahead.

It is possible to observe that the predicted values have the same behavior as the observed, especially in the predictions of the next hour, where the predicted values are closer to those observed.

It is interesting to note that, despite possible changes in the climatological scenario that occurred over the years, the model managed to capture the non-linear relationships between the various atmospheric parameters used for its training, being able to infer how the tropospheric ozone concentrations would behave in the test data. This shows how neural



networks, once well trained and generalized, are able to show good results, even in situations in which data and patterns never seen before are presented.

Fig. 10 shows the loss graph of the training and validation. It is possible to observe that there is a difference of about 0.01 of MSE, between the final epoch of the training and validation phases, thus showing that the model achieved a good generalization, with no under or overfitting.

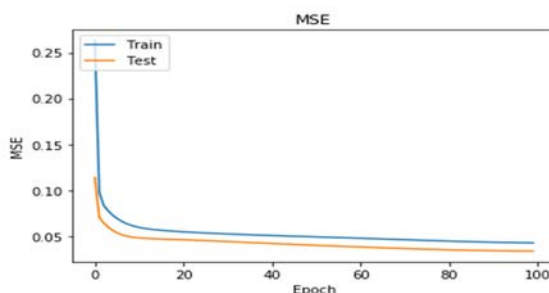


Figure 10: LSTM model training and validation loss graph.

## 6 CONCLUSIONS

The present work presented the great potential of predicting tropospheric ozone levels in an urban area with the use of the long short-term memory recurrent neural network, especially for long-term forecasts. The neural network managed to capture the non-linear relationships between the various atmospheric parameters used for its training and it was able to infer how the tropospheric ozone concentrations would behave in the test data. Health authorities can use these kinds of deep learning prediction models as an aid to prevent health problems due to exposure to the pollutant. As future work, the model developed will be validated over other air monitoring stations in other cities to assess capability to generalize and predict ozone levels in urban areas. An ensemble model or a hybrid system that combines the power of MLP along with LSTM models can also be developed and validated in order to explore the great potential for the prediction of the short to medium-term tropospheric ozone concentrations.

## ACKNOWLEDGEMENT

This work was developed at the Supercomputing Center for Industrial Innovation at SENAI CIMATEC.

## REFERENCES

- [1] World Health Organization, Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide: Report on a WHO working group, World Health Organization, 2003.
- [2] Monks, P.S. et al., Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. *Atmospheric Chemistry and Physics Discussions*, **14**(23), 2015.
- [3] Hogsett, W.E., Weber, J.E., Tingey, D., Herstrom, A., Lee, E.H. & Laurence, J.A., Environmental auditing: An approach for characterizing tropospheric ozone risk to forests. *Environmental Management*, **21**(1), pp. 105–120, 1997.



- [4] World Health Organization, *Occupational and Environmental Health Team, Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide, and Sulfur Dioxide: Global Update 2005: Summary of Risk Assessment*, World Health Organization, 2006.
- [5] Sousa, S.I.V., Martins, F.G., Alvim-Ferraz, M.C.M. & Pereira, M.C., Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling and Software*, **22**(1), pp. 97–103, 2007.
- [6] Gomez-Sanchis, J., Martín-Guerrero, J.D., Soria-Olivas, E., Vila-Francés, J., Carrasco, J.L. & del Valle-Tascón, S., Neural networks for analysing the relevance of input variables in the prediction of tropospheric ozone concentration. *Atmospheric Environment*, **40**(32), pp. 6173–6180, 2006.
- [7] Alves, L., Nascimento, E.G.S. & Moreira, D.M., Hourly tropospheric ozone concentration forecasting using deep learning. *WIT Transactions on Ecology and the Environment*, vol. 236, WIT Press: Southampton and Boston, pp. 129–138, 2019.
- [8] Biancofiore, F. et al., Analysis of surface ozone using a recurrent neural network. *Science of the Total Environment*, **514**, pp. 379–387, 2015.
- [9] Elman, J.L., Finding structure in time. *Cognitive Science*, **14**(2), pp. 179–211, 1990.
- [10] Pak, U., Kim, C., Ryu, U., Sok, K. & Pak, S., A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction. *Air Quality, Atmosphere and Health*, **11**(8), pp. 883–895, 2018.
- [11] Hochreiter, S. & Schmidhuber, J., Long short-term memory. *Neural Computation*, **9**(8), pp. 1735–1780, 1997.
- [12] Cho, K. et al., Learning phrase representations using RNN encoder–decoder for statistical machine translation, 2014. arXiv preprint, arXiv:1406.1078.
- [13] Chung, J., Gulcehre, C., Cho, K. & Bengio, Y., Empirical evaluation of gated recurrent neural networks on sequence modelling, 2014. arXiv preprint, arXiv:1412.3555.

