**Journal of Atomic, Molecular, Condensed Matter & Nano Physics**
Vol. 7, No. 3, pp. 197–206, 2020
ISSN 2582-8215 (online)
Published by RGN Publications
DOI: 10.26713/jamcnp.v7i3.1545

http://www.rgnpublications.com

**Proceedings of ICACTCE'21**
High School of Technology, Moulay Ismail University Meknes, Morocco, and
Faculty of Sciences and Techniques Mohammedia, Hassan II University, Morocco
March 24 – 26, 2021, Morocco
***Editors***: Mariyam Ouaissa, Mariya Ouaissa, Sarah El Himer, and Zakaria Boulouard

Research Article

# The Use of a Recurrent Neural Network for Forecasting Ozone Concentrations in the City of Agadir (Morocco)

Anas Adnane *[1], Redouane Leghrib[1], Jamal Chaoufi[1] and Ahmed Chirmata[2]

[1] *LETSMP, Department of Physics, Faculty of Science, Ibn Zohr University, Agadir, Morocco*
[2] *Department of Energy and Environment, Wilaya of Agadir, Agadir, Morocco*

**Abstract.** Air quality is a complex issue which depends mutually on source emission, land topography, meteorological parameters and used mathematical tools for forecasting its dispersion. One of major toxic gas is ozone which could be dangerous to human health. The present work has been developed to forecast ozone concentrations in the city of Agadir using a *Recurrent Neural Network* (RNN). Predicting ozone concentrations will provide useful information, especially, for decision makers in order to prevent and reduce the ozone human health impacts. The data was collected in the most polluted area in the city using a mobile monitoring station during a period of 60 days. We have tested different neural network architectures and we found that the 1-hour forecast model, whose input parameters are a combination of meteorological parameters as well as CO and $SO_2$, give the most optimal results. *Coefficient of Correlation* (CC), *Root Mean Square Error* (RMSE) and *Mean Absolute Error* (MAE) were used to evaluate the statistical agreement between observed and predicted values. The model successfully predicts the ozone concentration by a bias of 4 $\mu$g.m$^{-3}$ over 24 hours and which a correlation coefficient is more than 80%. This work highlights the ability of the recurrent neural networks to forecast air pollutant concentrations in urban areas.

**Keywords.** Recurrent neural network; Machine learning; Pollution; Air quality; Ozone; Morocco

**PACS.** 84.35.+i; 92.60.Sz; 07.05.Tp



*[*] **Corresponding author:** anas.adnane@edu.uiz.ac.ma

## 1. Introduction

Ozone is a secondary pollutant resulting from photochemical processes involving nitrogen oxides and volatile organic compounds in the lower atmosphere [12]. At ground level, ozone is considered as a noxious gas. However, in the upper atmosphere, ozone forms a layer that shields the Earth from ultraviolet radiation. Because of their negative impacts on human health and the environment, predicting air pollutant concentrations and especially those of ozone has become one of the objectives of scientists and decision-makers. Ozone can cause negative damage to human health, such as increased frequency of asthma attacks, difficulty in breathing deeply or even damage to the lungs [9]. The forecast can therefore provide legislators an idea about the future situation so that they can take measures that reduce these concentrations. This information is also useful for the population, especially vulnerable people, such as the elderly, asthmatics and children, to avoid certain affected areas [7].

Some meteorological parameters have an influence, direct or indirect, on concentration evolution of ozone in the atmosphere. The best known are the temperature and solar radiation, at high values, leading to higher concentrations of ozone [16]. Wind also has an impact on ozone concentrations: increasing wind speed means decreasing ozone concentrations. Thus, wind direction and speed influence the ambient ozone concentration at a fixed position [14]. Relative humidity and precipitation, on the other hand, lead to lower concentrations of ozone. This is explained by the reduction in the efficiency of photochemical production [11].

*Multiple Linear Regression* (MLR) is one of the most widely used methods for modeling ozone concentrations. This method models the linear relationship between the concentrations of $O_3$ and other parameters. In contrast, the relationship between ozone and other variables is non-linear and complex [6]. For this reason, machine learning algorithms in general and *Artificial Neural Networks* (ANNs) in particular, provide a more appropriate alternative for modeling nonlinear relationships [15]. However, ANNs find it difficult to predict high or extreme levels of pollutant concentrations [1, 2, 4]. Despite their advantages, few studies have been done concerning on modeling of ambient ozone by using neural networks [13], and especially rare works have been developed concerning Moroccan air quality forecasting.

In this work, we have established and tested several types of recurrent neural networks taking into account the number of hidden layers and also the number of neurons per layer. Furthermore, each structure was evaluated by varying the combination between metrological parameters and some primary pollutants (CO and $SO_2$). These different architecture leads to predict ozone concentrations over the city of Agadir. The obtained results were evaluated by means of *Coefficient of Correlation* (CC), *Root Mean Square Error* (RMSE) and *Mean Absolute Error* (MAE). Two forecasting duration were taking into consideration (1-h and 24-h in advance). The present work was organized as follow: we started with the definition of studied area, and giving standard statistical information about the used air quality database concerning metrological and pollutants emissions, then we focused on the development of Non-linear Autoregressive Neural Network with Multiple Exogenous Variables for ozone forecasting by testing different architectures and varying the inputs of the model. Finely, the performance of the model was evaluated based on statical parameters such as CC, RMSE and MAE.

# 2. Site and Observations

## A. Study Area

Agadir is a city located in the southwest of the Moroccan kingdom. It is characterized by a semi-arid climate. Agadir is the capital of the prefecture of Agadir-Ida-Outanane. This prefecture covers an area of 2,297 km$^2$ with a population of 506,517.

Since its reconstruction after the 1960 earthquake, Agadir has become an international seaside tourist destination. Besides tourism, the economy of this city is mainly based on agriculture and fishing.

## B. Database

Our observation site is in Anza (−9.658485;30.448091), which is one of the most polluted areas in the city of Agadir. A series of hourly measurements of meteorological parameters (temperature, wind direction, wind speed and humidity) and of some pollutants ($O_3$, CO, $SO_2$) was carried out by Chirmata *et al*. [5]. We have chosen the period [05/01/2016-06/30/2016] since it contains fewer missing data. The selection of pollution variables as well as meteorological ones was based on their availability. A statistical summary of the variables measured in the study period at the Anza station is shown in Table 1.

**Table 1.** Characteristics of the variables measured at the Anza site

| Parameter | Unity | Range [Min ;Max] | Average | %Missing data | Standard deviation |
|---|---|---|---|---|---|
| $O_3$ | µg/m$^3$ | [2,84 ; 89,70] | 48,76 | 5,51 | 15,07 |
| Temperature | °C | [18,53 ; 28,11] | 22,92 | 5,79 | 0,86 |
| Humidity | % | [0,26 ; 90,14] | 35,12 | 5,51 | 41,20 |
| Wind speed | m/s | [0,00 ; 0,96] | 0,09 | 5,51 | 0,09 |
| Wind direction | ° | [0,03 ; 358,74] | 214,17 | 5,51 | 72,42 |
| $SO_2$ | µg/m$^3$ | [0,00 ; 98,77] | 13,78 | 5,51 | 19,45 |
| CO | µg/m$^3$ | [0,00 ; 2,70] | 0,51 | 5,93 | 0,19 |

## C. Data Analysis

### a. Missing Data

**Table 2.** The different RNN models used and their characteristics

| RNN | Timestep | Input parameters |
|---|---|---|
| M1 | 1 hour | Meteo |
| M2 | 1 hour | Meteo + CO + $SO_2$ |
| M3 | 24 hours | Meteo |
| M4 | 24 hours | Meteo + CO + $SO_2$ |

Missing data is a classic problem for the scientific community. This problem requires reconstruction in order to move forward. These missing data can be estimated using several methods. In this work, we have used the nearest neighbor method. It consists of replacing the missing data by the average of data from days $d-1$ and $d+1$.

**b. Zonal & Meridional Wind**

We decided to transform the data related to the wind into two components. Wind $X$ represents the component of the southerly wind or also called zonal wind and Wind $Y$ indicates the component of the easterly wind which is no other than the meridional wind. This transformation allows us to take into account the cyclical characteristics of the wind and to avoid sudden jumps in values. The formulas used are as follows:

$$\text{Wind } X = w * \cos(v), \tag{2.1}$$

$$\text{Wind } Y = -w * \sin(v), \tag{2.2}$$

where $w$ indicates wind speed and $v$ represents wind direction.

## 3. Ozone Forecasting Models

**A. NARX (Non-linear Autoregressive Neural Network with Multiple Exogenous Variables)**

Among the tools most used in nonlinear modeling applications, we find the NARX-NN (Nonlinear Autoregressive with Exogenous Input – Neural Network) models, also called RNN (Recursive Neural Network). This model is described as a recursive discrete-time input-output equation, which is written as:

$$\tilde{y}(t) = F[x(t-1), \dots, x(t-n_x), y(t-1), \dots, y(t-n_y)], \tag{3.1}$$

where $n_x$ and $n_y$ denote the maximum lags of the exogenous and endogenous variables $x$ and $y$, respectively, and $y(t)$ indicates the one-step forecast of the actual value $\tilde{y}(t)$. The function $F$ is represented by a *Feed Forward Multi-Layer Perceptron* (FF-MLP). RNNs are therefore only an extension of FF-MLP.

Unlike feed-forward ANNs, as Figure 1 shows, a *Recurrent Neural Network* (RNN) provides feedback. In other words, some output neurons are connected to neurons from previous layers, which can improve the learning capacity of RNNs and give better results [8, 10].

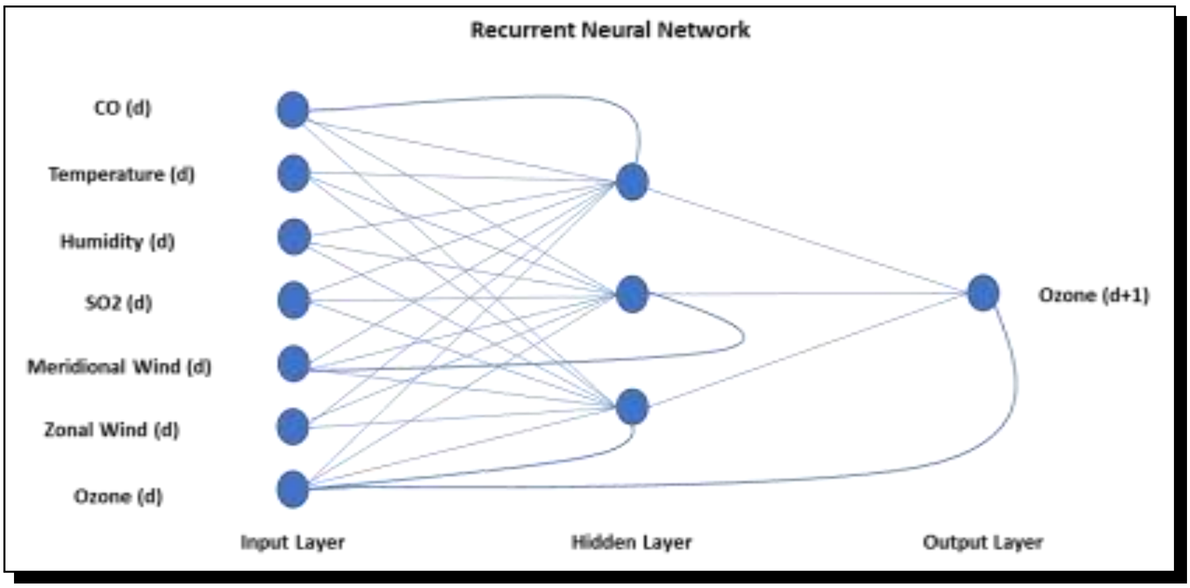


**Figure 1.** General scheme of a Recurrent Neural Network (RNN)

The most well-known topology of an RNN is the Elman network [17]. This is typically a multi-layered network with feedback from the last hidden layer output to the first hidden layer input. These connections allow Elman's network to take into account the time parameter.

ANNs consist of an input layer, an output layer, and one or more hidden layers, each having a varying number of interconnected neurons. The input layer receives the input data. As for the hidden layers, they allow adjustment and minimize the differences between input and output. The challenge for neural network users is to find the optimal size (number of hidden layers and number of neurons) as well as the best activation function for each application [3,5]. The output layer represents the parameter that we want to model, calculated by the following formula:

$$y_i = f\left(\sum_{k=1}^{N} x_k w_{k,i} + b_i\right), \tag{3.2}$$

where $N$ is the number of nodes in the previous layer, $x$ represents data from hidden layers, $w$ represents weight, $b$ is bias, and $y$ represents the output parameter. The learning process is performed repeatedly in a gradient descent algorithm until a stopping criterion is satisfied.

**Table 3.** Performance results of different models for the hourly prediction of ozone where the input parameters are meteorological variables

| Timestep | Structure | CC (%) | MAE | RMSE |
|---|---|---|---|---|
| 1 | 12 | 45,27 | 7,35 | 8,97 |
| 1 | 10 | 65,75 | 8,87 | 10,70 |
| 1 | 8 | **77,55** | 7,46 | 8,81 |
| 1 | 6 | 74,13 | 5,79 | 7,26 |
| 1 | 12-12 | 31,25 | 9,61 | 11,44 |
| 1 | 10-10 | 64,19 | 10,81 | 12,84 |
| 1 | **8-8** | 73,63 | **4,95** | **6,03** |
| 1 | 6-6 | 69,44 | 5,86 | 7,52 |
| 24 | 12 | 44,46 | 6,69 | 7,91 |
| 24 | 10 | 36,20 | 10,54 | 13,06 |
| 24 | 8 | **56,87** | 7,90 | 10,45 |
| 24 | 6 | 34,53 | 11,13 | 13,46 |
| 24 | 12-12 | 53,43 | 7,62 | 9,57 |
| 24 | 10-10 | 54,46 | 7,31 | 9,05 |
| 24 | **8-8** | 55,01 | **6,36** | **7,36** |
| 24 | 6-6 | 47,25 | 7,02 | 9,17 |

### B. Model Development

Finding the structure of the neural network that gives the most optimal results is a complex and difficult task. This difficulty lies in the fact that no rule exists for determining the number of neurons, the number of hidden layers, or even the activation function. Therefore, we tested several models in order to find the most optimal and adequate one taking into account the correlation coefficient and obtained bias to the observation. We tested four types of models: a model whose input parameters are meteorological variables only (temperature, humidity, zonal wind and southerly wind) and ozone, and the second by adding two pollutants (CO and $SO_2$). The other two models are similar to the two previous models but with a 24-hour time step instead of an hour.

Then we divided our database into two categories for a total of 1416 times sample: 85% of the data, 1204 cases for training the neural network and 15% of the data, 212 cases for evaluation and validation of the quality and performance of the model.

Table 2 shows the different structures of the neural networks that we tested. In most cases, we have resorted to the use of two hidden layers in order to optimize the results. The logistic and linear sigmoid activation functions were used in the hidden layers and the output layer respectively. For the training of the RNNs, we used the Levenberg-Marquardt backpropagation algorithm. We programmed our algorithms using Matlab R2015a.

**Table 4.** Performance results of different models for the hourly prediction of ozone where the input parameters are meteorological variables as well as $SO_2$ and CO

| Timestep | Structure | CC (%) | MAE | RMSE |
|---|---|---|---|---|
| 1 | 12 | 74,97 | 5,52 | 4,54 |
| 1 | **10** | **83,37** | 4,77 | **3,71** |
| 1 | 8 | 82,44 | **4,69** | 3,82 |
| 1 | 6 | 82,44 | 4,74 | 3,78 |
| 1 | 12-12 | 79,41 | 5,80 | 4,83 |
| 1 | 10-10 | 78,79 | 5,13 | 4,33 |
| 1 | 8-8 | 80,69 | 4,94 | 4,13 |
| 1 | 6-6 | 79,63 | 5,39 | 4,42 |
| 24 | 12 | 44,15 | 12,56 | 1,33 |
| 24 | 10 | 34,42 | 14,56 | 11,51 |
| 24 | 8 | 42,76 | 8,83 | 7,02 |
| 24 | 6 | 47,53 | 10,53 | 8,45 |
| 24 | 12-12 | 37,87 | 8,53 | 6,76 |
| 24 | 10-10 | 55,44 | 10,23 | 8,46 |
| 24 | **8-8** | **74,70** | **7,63** | **6,56** |
| 24 | 6-6 | 52,38 | 8,64 | 6,63 |

To facilitate the nomenclature of the models, we will name M1 the 1-h ahead forecast model whose input parameters are only the meteorological variables. M2 will be the forecast model 1-h ahead and whose input parameters are the meteorological variables as well as the pollutants. M3 and M4 will be similar to M1 and M2, respectively, but with a 24-h ahead forecast.

## 4. Results and Discussion

### A. Model Evaluation (CC, MAE, RMSE)

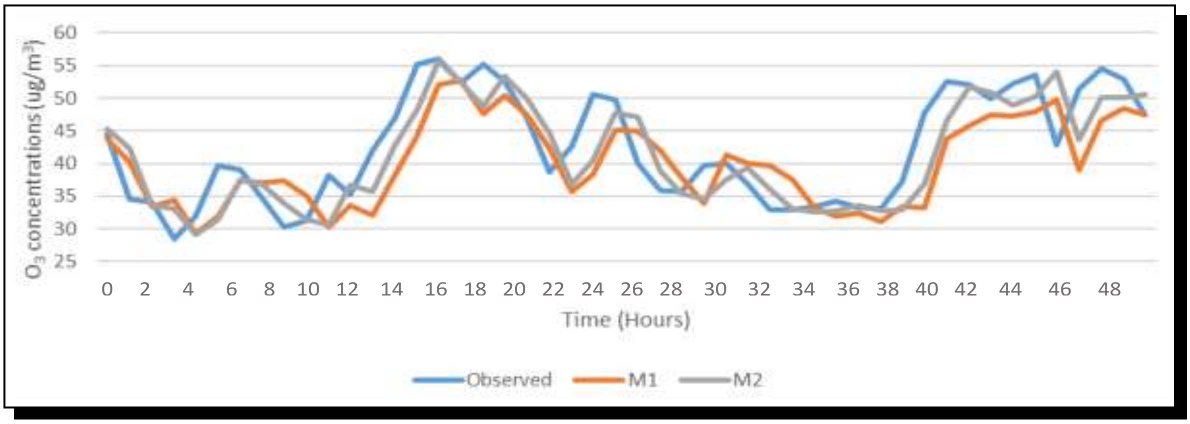


**Figure 2.** Comparison of observed values of ozone concentrations and those predicted by NARX-ANN models 1-h in advance

In order to assess the performance of the methods used for the prediction of ozone concentrations, three statistical parameters were calculated, including the *Mean Absolute Error* (MAE), the *Random Mean Square Error* (RMSE) and the *Correlation Coefficient* (CC). The results with the highest CC value and the lowest value of RMSE as well as MAE are the most favorable. These criteria were calculated using the following formulas:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}, \tag{4.1}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{2}\sum_{i=1}^{n}(y_i - x_i)^2}, \tag{4.2}$$

$$\mathrm{CC} = \frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)(x_i - \bar{x}_i)}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2(x_i - \bar{x}_i)}}, \tag{4.3}$$

where $x_i$ and $y_i$ represent respectively the values of the predicted and observed values, and $\bar{x}_i$ and $\bar{y}_i$ are respectively the mean values of the predicted and observed values.

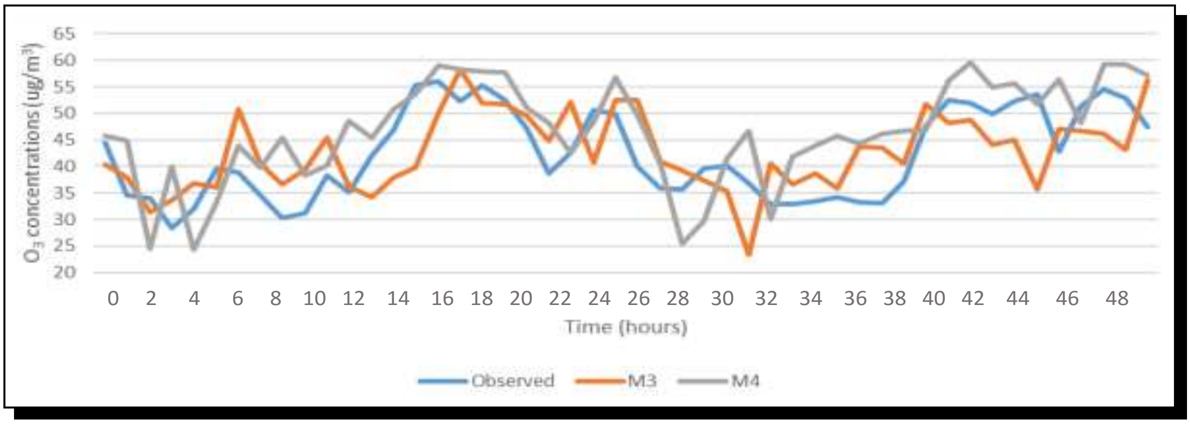


**Figure 3.** Comparison of observed values of ozone concentrations and those predicted by NARX-ANN models 24-h in advance

### B. Simulation Results and Model Performance

We denote by a structure $x$, a network of neurons with a hidden layer whose number of neurons is $x$. A $y$-$y$ structure in turn refers to a neural network with two hidden layers whose number of neurons in each of them is $y$.

From the results shown in Table 3 and Table 4, structure 8-8 gives the best results for all three model types M1, M3, and M4. As for the M2 model, the most optimal results were obtained with structure 10. Table 5 showcase the best performing models.

By comparing the 1-h ahead forecast models, we notice that M2 performs the best with a CC of 0.83, an RMSE of 4.77 and an MAE of 3.71. This highlights the added value of emissions in predicting ozone. Figure 2 combines the observations of ozone concentrations for the first 48 hours of the test period and the concentrations predicted by models M1 and M2. Note that the estimates of the two models are very close to the real values with a slightly better estimate for the M2 model. The models were able to estimate the extreme values well.

**Table 5.** Performance results of the best performing models for the hourly prediction of ozone

| Model | Structure | CC (%) | RMSE | MAE |
|---|---|---|---|---|
| M1 | 8-8 | 73,63 | 4,95 | 6,03 |
| M2 | 10 | 83,37 | 4,77 | 3,71 |
| M3 | 8-8 | 55,10 | 6,36 | 7,36 |
| M4 | 8-8 | 74,70 | 7,63 | 6,56 |

By comparing the 24-hour forecast models, we can see that M4 performs the best with a CC of 0.75, RMSE of 7.63 and MAE of 6.56. This once again confirms the importance of emissions as predictors of ozone. While meteorological variables are essential for forecasting ozone, adding emissions as predictors allows forecasts to be even closer to observations. In addition, as expected, we notice that the model results decreased when moving from a one hour forecast to a 24-hour forecast. Figure 3 combines the observations of ozone concentrations for the first 48 hours of the test period and the concentrations predicted by models M3 and M4. Note that the two models fail to well predict extreme values. But in general, the results obtained are close to reality with sometimes an overestimation or even an underestimation.

The results obtained show how efficient RNNs are. Despite the use of few input parameters and a more or less average data period, Recurrent neural networks were able to provide very satisfactory results. Note that little work has been done in Africa in general and in Morocco in particular, to model ambient ozone concentrations using Machine Learning methods.

## 5. Conclusion

In this work, we have developed models based on recursive neural networks to obtain forecasts of ozone concentrations at 1-hour and 24-hour ahead of the city of Agadir. The comparison between the models shows that the most optimal results were obtained by the 1-hour forecast model, whose input parameters are a combination of meteorological parameters as well as some pollutants, namely CO and $SO_2$. The addition of pollutants in the input parameters turned out to be more relevant, thus making the model more efficient.

The use of NARX models for predicting pollutant concentrations in general and ozone in particular, is promising. The results obtained are very satisfactory despite the limitation of the data. The methodology used in this work can be generalized, but the models used are specific for the city of Agadir and the used database.

In view of this work, more work is ongoing on the adaptation of this method to model and forecast other air pollutants over the city of Agadir. Other aspect such as chemical reactions, land use etc should be taking into our model that could further optimize its outcomes.

## Acknowledgment

We acknowledge the Wilaya and the Region of the Souss Massa for their collaboration and supporting the Research, under the leadership of the Wali and the President of Souss Massa Region, to whom we present our great and special thanks.

### Competing Interests

The authors declare that they have no competing interests.

### Authors' Contributions

All the authors contributed significantly in writing this article. The authors read and approved the final manuscript.

## References

**[1]** Y. Bai, B. Zeng, C. Li and J. Zhang, An ensemble long short-term memory neural network for hourly PM 2.5 concentration forecasting, *Chemosphere* **222** (2019), 286 – 294, DOI: 10.1016/j.chemosphere.2019.01.121.

**[2]** M. Banja, D.K. Papanastasiou, A. Poupkou and D. Melas, Development of a short–term ozone prediction tool in Tirana area based on meteorological variables, *Atmospheric Pollution Research* **3** (2012), 32 – 38, DOI: 10.5094/APR.2012.002.

**[3]** S.V. Belavadi, S. Rajagopal, R. Ranjani and R. Mohan, Air quality forecasting using LSTM RNN and wireless sensor networks, *Procedia Computer Science* **170** (2020), 241 – 248, DOI: 10.1016/j.procs.2020.03.036.

**[4]** M. Catalano, F. Galatioto, M. Bell, A. Namdeo and A.S. Bergantino, Improving the prediction of air pollution peak episodes generated by urban transport networks, *Environmental Science & Policy* **60** (2016), 69 – 83, DOI: 10.1016/j.envsci.2016.03.008.

**[5]** A. Chirmata, R. Leghrib and I.A. Ichou, Implementation of the air quality monitoring network at Agadir city in Morocco, *Journal of Environmental Protection* **8** (2017), 540 – 567, DOI: 10.4236/jep.2017.84037.

**[6]** A.C. Comrie, Comparing neural networks and regression models for ozone forecasting, *Journal of the Air & Waste Management Association* **47** (1997), 653 – 633, DOI: 10.1080/10473289.1997.10463925.

**[7]** D. Dunca, A. Pohoata and S. Lordache, Using wavelet-feedforward neural networks to improve air pollution forecastion in urban environments, *Environmental Monitoring and Assessment* **187**(7) (2015), 477, DOI: 10.1007/s10661-015-4697-x.

**[8]** J.L. Elman, Finding structure in time, *Cognitive Science* **14** (1990), 179 – 211, DOI: 10.1207/s15516709cog1402_1.

**[9]** EPA, *Health Effects of Ozone Pollution*, United States Environmental Protection Agency, `https://www.epa.gov/ground-level-ozone-pollution/health-effects-ozone-pollution`.

**[10]** M.T. Hagan, H.B. Demuth and M.H. Beale, *Neural Network Design*, 1st edition, PWS Publishing Company, Boston, p. 736 (1996).

**[11]** J. Lelieveld and P.J. Crutzen, Influence of cloud and photochemical processes on tropospheric ozone, *Nature* **343** (1990), 227 – 233, DOI: 10.1038/343227a0.

**[12]** S.C. Liu, M. Trainer, F.C. Fehsenfeld, D.D. Parrish, E.J. Williams, D.W. Fahey, G. Hübler, P.C. Murphy, Ozone production in the rural troposphere and the implications for regional and global ozone distributions, *Journal of Geophysical Research: Atmospheres* **92** (1987), 4191 – 4207, DOI: 10.1029/JD092iD04p04191.

**[13]** R. Ma, J. Ban, Q. Wang and T. Li, Statistical spatial-temporal modeling of ambient ozone exposure for environmental epidemiology studies: a review, *Science of the Total Environment* (2019), DOI: 10.1016/j.scitotenv.2019.134463.

**[14]** J. Mao, L. Wang, C. Lu, J. Liu, M. Li, G. Tang, D. Ji, N. Zhang and Y. Wang, Meteorological mechanism for a large-sclae persistent severe ozone pollution event over eastern China in 2017, *Journal of Environmental Sciences* **92** (2020), 187 – 199, DOI: 10.1016/j.jes.2020.02.019.

**[15]** X. Ren, Z. Mi and P.G. Georgopoulos, Comparison of machine learning and land use regression for fine scale spatiotemporal estimation of ambient air pollution: modeling ozone concentrations across the contiguous United States, *Environment International* **142** (2020), 105827, DOI: 10.1016/j.envint.2020.105827.

**[16]** P. Zanis, P. Hadjinicolaou, A. Pozzer, E. Tyrlis, S. Dafka, N. Mihalopoulos and J. Lelieveld, Summertime free-tropospheric ozone pool over the eastern Mediterranean/Middle East, *Atmospheric Chemistry and Physics* **14** (2014), 115 – 132, DOI: 10.5194/acp-14-115-2014.

**[17]** J. Zhang and W. Ding, Prediction of air pollutants concentrations based on an extreme learning machine: the case of Hong Kong, *International Journal of Environmental Research and Public Health* **14**(2) (2017), 114, DOI: 10.3390/ijerph14020114.