COMPUTER SCIENCE

Unorganized machines and linear multivariate regression model applied to atmospheric pollutant forecasting

Daniel Silva Campos[1], Yara de Souza Tadano[2*], Thiago Antonini Alves[2], Hugo Valadares Siqueira[2] and Manoel Henrique de Nóbrega Marinho[1]

[1]Escola Politécnica de Pernambuco, Universidade de Pernambuco, Rua Benfica, 455, 50720-001, Recife, Pernambuco, Brazil. [2]Universidade Tecnológica Federal do Paraná, Rua Doutor Washington Subtil Chueire, 330, 84017-220, Ponta Grossa, Paraná, Brazil. *Author for correspondence. E-mail: yaratadano@utfpr.edu.br

ABSTRACT. Air pollution is a relevant issue studied worldwide, and its prediction is important for social and economic management. Linear multivariate regression models (LMR) and artificial neural networks (ANN) are widely applied to forecasting concentrations of pollutants. However, unorganized machines are scarcely used. The present investigation proposes the application of unorganized machines (echo state networks - ESN and extreme learning machines - ELM) to forecast hourly concentrations of particulate matter with the aerodynamic diameter up to 10 μm (PM_{10}), carbon monoxide (CO), and ozone (O_3) at the metropolitan region of Recife, Pernambuco, Brazil. The results were compared with multilayer perceptron neural network (MLP) and LMR. The prediction was made using or not meteorological variables (wind speed, temperature, and relative humidity) as input data. The results showed that the inclusion of these variables could increase the general performance of the models considering one step ahead forecasting horizons. Also, the ELM and the LMR achieved the best overall results.

Keywords: forecasting; air pollutants; linear multivariate; regression models; echo state networks; extreme learning machines.

Introduction

Air pollution is a widely studied subject mainly due to its impact on human health. It is considered a major environmental risk to public health by the World Health Organization (WHO), which estimates around eight million deaths per year due to the exposure to air pollution (Jasarevic & Lindmeier, 2015). The association between concentrations of air pollution and severe health problems is a well-studied subject worldwide, as we can see in Langrish et al. (2012), Wu et al. (2012), Tadano, Siqueira, and Alves (2016), Polezer et al. (2018), Ardiles et al. (2018), Kachba et al. (2020).

Bad air quality causes chronic diseases and premature mortalities (Cabaneros, Calautit, & Hughes, 2019). In this sense, it is very important to the policy-makers and urban city planners to have previous information about air quality in order to find solutions to avoid or minimize the effects of air pollution on human health.

Therefore, air quality management is essential to improve the quality of life in urban areas. One major problem which aggravates air quality of these areas are the poorly designed building structures, allied to a large population density. The possibility to predict episodes of critical high atmospheric pollution a few days in advance would be an efficient way to help authorities take preventive and evasive action and protect citizen health (Nagendra & Khare, 2006; Caselli, Trizio, Gennaro, & Ielpo, 2009; Moustris, Ziomas, & Paliatsos, 2010).

The most commonly used tools for air quality analysis and forecasting are statistical regression models. Researchers have recently been applying artificial neural networks (ANN) and comparing their performance with classical statistical models (Baawain & Al-Serihi, 2014; Taylor, Retalis, & Flocas, 2016). To improve the prediction of the air pollutants concentration, some researchers included meteorological variables (Moustris, Larissi, Nastos, Koukouletsos, & Paliatsos, 2013).

Grivas and Chaloulakou (2006) used an optimization procedure by genetic algorithm to select variables, resulting in a better performance than when using developed multiple linear regression models.

Stadlobera, Hörmannb, and Pfeiler (2008) used linear regression models, combining daily pollutant concentrations with the meteorological variables of previous days (wind speed, precipitation, and

temperature), aiming to model the daily mean concentrations of particulate matter with aerodynamic diameter up to 10 μm (PM_{10}). The results showed that this approach supported future forecasting of pollutants.

Caselli et al. (2009) used neural networks (multilayer perceptron - MLP and radial basis function - RBF) and a multivariate regression model (MRM) to predict daily PM_{10} concentrations, considering meteorological information (temperature, wind speed, pressure, and relative humidity) as input data for one, two, and three days in advance. The authors concluded that ANN showed more accurate results than the multivariate regression model, mainly for one day forecasting. They also highlighted that MRM failed to predict high PM_{10} values.

Moustris et al. (2010) showed that MLP has a good capacity to forecast the air quality of three consecutive days. They forecasted maximum daily values of the European Regional Pollution Index (ERPI) for CO, NO_2, SO_2, and O_3 three days in advance using meteorological variables such as air temperature, relative humidity, and wind speed and direction. They also predicted the number of hours during the day with the concentration of at least one of the pollutants above the threshold.

Moustris, Nastos, Larissi, and Paliatsos (2012) predicted 24 hours ozone concentrations in Athens using multiple linear regression and the MLP with wind speed, temperature, and ozone concentrations as explanatory variables. They concluded that the ANN performed better.

Moustris et al. (2013) used a forecasting of 24 hours in advance for the daily concentrations of PM_{10} and the number of hours exceeding PM_{10} concentration threshold during the day in five regions within the greater Athens area. The authors used MLP and considered maximum hourly NO_2 concentration, air temperature, relative humidity, and wind speed and direction as input variables. They concluded that the ANN can successfully predict daily threshold exceedances and can be used to forecast the concentration of PM_{10} one day in advance.

Sayegh, Munir, and Habeebullah (2014) studied different approaches, such as linear and nonlinear regressions, and learning machine methods to evaluate which gives a more accurate prediction of PM_{10} concentration in urban areas. They used meteorological variables (relative humidity, temperature, and wind speed) and the concentration of other pollutants (CO, SO_2, and NO_x). The results showed that multiple linear regression and quantile regression were the best models to describe PM_{10} variation.

Mattos Neto, Madeiro, Ferreira, and Cavalcanti (2014) developed neural network architecture such as multilayer perceptron (MLP) with parameter optimization by genetic algorithm and tested on the prediction of PM_{10} and particulate matter with aerodynamic diameter up to 2.5 μm ($PM_{2.5}$). The experimental study considered the random behavior of pollutant concentration. The results showed a consistent improvement when compared to other prediction techniques.

Zhang and Ding (2017) proposed to predict the concentration of air pollutants using extreme learning machine (ELM) with meteorological parameters (temperature, relative humidity, and wind speed and direction) and the concentrations of NO_2, NO_x, O_3, $PM_{2.5}$, and SO_2 in Hong Kong. The results showed ELM with slightly better performance than multiple linear regression and MLP.

Some recent studies applied ensembles to increase the quality of pollutant forecasting (Siwek & Osowski, 2012; Firmino, Mattos Neto, & Ferreira, 2014; Gong & Ordieres-Meré, 2016; Wang & Song, 2018; Bai, Zeng, Li, & Zhang, 2019). This methodology combines the outputs of various models, which were pre-tuned. The authors usually use a neural network, mainly the MLP, as combiner to increase the mapping capability of single models. As discussed by Bai et al. (2019), this approach can be considered state of the art.

However, according to the available literature, a few researchers used unorganized machines to predict concentrations of air pollution (Tadano, Siqueira, Alves, & Marinho, 2017; Zhang & Ding, 2017). Recent researches have addressed the echo state networks (ESN) and ELM to problems relating air quality and health (Polezer et al., 2018; Araujo, Belotti, Alves, Tadano, & Siqueira, 2020; Kachba et al., 2020). The use of such architectures is relevant since they have desirable characteristics to any forecasting problems: easiness in understanding and implementation and fast convergence allied to good performances. These works showed that the structures could overcome important models, such as the MLP, Radial Basis Function Networks (RBF), and Generalized Linear Models (GLM). Furthermore, some linear approaches are not capable of forecasting some series, especially those in which there are missing values, a reality in developing countries like Brazil.

We highlight that the use of single neural models must be evaluated since the computational cost to adjust their free coefficients is lower than the ensembles which requires the adjustment of the single models and the combiner.

Therefore, this research aims to compare the performance of the unorganized machines (ELM and ESN) with MLP and traditional linear multivariate regression modeling to forecast air pollution (PM_{10}, CO, and O_3) in the metropolitan region of Recife, Pernambuco, Brazil. We compared the performance using and not using meteorological variables (relative humidity, temperature, and wind speed) to identify if meteorological variables improve the forecasting of pollutant concentrations.

The rest of the paper is organized in five sections, in which we discuss the databases addressed, the LMR model, the neural networks used, the computational results, as well as the critical analysis, and the conclusions.

Materials and methods

The database was obtained by the Environmental Agency of Pernambuco (CPRH), consisting of concentrations of PM_{10} [$\mu g.m^{-3}$], carbon monoxide (CO) [ppm], and ozone (O_3) [$\mu g.m^{-3}$], and the meteorological variables relative humidity (RH) [%], average temperature (AT) [°C], and wind speed (WS) [$m.s^{-1}$], from July 17th, 2015 to July 17th, 2017. All data comprise hourly measures from the air quality monitoring station (called *Estação Gaibú*) installed at the *Frei Otto*, *Nossa Senhora do Ó* Highschool, located at Ipojuca, Pernambuco, Brazil. Figure 1 shows the location of the Gaibú station on the right side, with coordinates 8º 19' 44. 37" S and 34º 57' 22. 66" W. On the left side down of Figure 1, the Abreu e Lima Refinery (RNEST) is highlighted at PE 60 Highway, which is one of the most significant sources of air pollution around the studied area.

The state of Pernambuco is inserted in an intertropical zone presenting predominantly high temperatures. In the region of Recife, for example, the average temperature is 25ºC, with a maximum of 32ºC. In inland cities, the temperature during winter months (between May and July) can drop considerably, reaching 8ºC in some places (*Governo do Estado de Pernambuco*, 2018). In the summer, the rainfall index reaches 515 mm on the coast, where the city of Ipojuca is located, while reaching 725 mm during the winter (*Agência Pernambucana de Águas e Clima* [APAC], 2018).

Table 1 shows the mean, standard deviation, and coefficient of variation to all input data. The coefficient of variation provides a variation measure of each data set regarding its mean to the studied period (July 17th, 2015, to July 17th, 2017). It is calculated by the ratio between the mean and the standard deviation. We may highlight that the mean concentration of PM_{10} is above the World Health Organization (WHO, 2006) guideline value of an 20 $\mu g\,m^{-3}$ annual mean and presents a higher variation during the year when compared to other pollutants (higher coefficient of variation).

We also obtained the correlation coefficients between all input data. Among the meteorological variables, relative humidity was the only one inversely proportional to the concentration of pollutants, showing correlation coefficient (R) with PM_{10}, CO, and O_3 of −0.560, −0.093, and −0.351, respectively. These results confirm that the higher the relative humidity, the lower the concentration of air pollutants, especially PM_{10}, due to wet deposition. It is important to highlight that, during winter, the relative humidity usually decreases, characterizing a dry season, which tends to increase respiratory diseases. The dry weather also increases the suspension of air pollutants, such as PM_{10}. Therefore, we can suppose that dryness, combined with higher concentrations of air pollutants may lead to an increase in respiratory diseases.

Linear multivariate regression model

Regression analysis is a statistical technique to model and investigate the relationship between two or more variables. A variety of engineering and science problems implicates in the exploration of this relationship (Montgomery & Runger, 2003). A mathematical model represents this kind of relationship, which is an equation associating the dependent variable (response variable - *Y*), which we desire to predict, with the independent variables (predictor or explanatory variables - x_j). Therefore, the regression analysis aims to uncover the behavior between the dependent variable and the independent ones. When the response variable *Y* depends on two or more independent variables, the relationship is given by Equation 1:

Figure 1. Location of Gaibú Air Quality Monitoring Station.

Table 1. Mean, standard deviation, and coefficient of variation of the concentration to pollutants and meteorological data from July 17th, 2015 to July 17th, 2017.

Parameter*	Mean	Standard Deviation	Coefficient of variation [%]
PM_{10} [µg.m^{-3}]	34.04	23.07	67.8
CO [ppm]	0.32	0.16	50.2
O_3 [µg.m^{-3}]	26.85	13.70	51.0
WS [m.s^{-1}]	3.34	1.06	31.6
RH [%]	74.34	9.43	12.7
AT [ºC]	27.80	2.08	7.5

*PM_{10}: particulate matter with aerodynamic diameter equal or less than 10 µm; CO: carbon monoxide; O_3: ozone; WS: wind speed; RH: relative humidity; AT: atmospheric temperature.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \qquad (1)$$

where:

β_0 represents the intersection, and the other parameters β_j, j = 1, 2, 3, ..., k are denominated the regression coefficients. The estimation of these parameters can be done by using the minimum square method.

Artificial neural networks

Artificial neural networks (ANN) have been, recently, widely applied and developed in atmospheric investigations due to their computational efficiency and generalization capability. A neural network is a simplified mathematical modeling of the natural neural system of the superior organisms. It is an effective alternative to solve times series forecasting instead of traditional linear models (Goyal & Kumar, 2011). The main advantage of this tool is the flexibility, which permits the wide application on several fields (Hooyberghs, Mensink, Dumont, Fierens, & Brasseur, 2005). Another advantage includes the ability to simulate nonlinear behaviors based on the input variables of the model (Dorling, Foxall, Mandic, & Cawley, 2003).

The multilayer perceptron (MLP) is the most used architecture to solve the kind of problem addressed in this research. Notwithstanding, the literature presents many other proposals such as Elman recurrent neural networks, pruned neural networks, radial basis, and others (Shahraiyni & Sodoudi, 2016). In this research, we considered the traditional MLP and two proposals of unorganized machines: extreme learning machine (ELM) and echo state network (ESN).

Feedforward neural networks (FNN) are those in which the information flows from the input layer to the output layer, flowing only in one direction. The main characteristic of these models is the universal

approximation capability that allows the neural models to approximate any nonlinear, continuous, limited, and differentiable function (Haykin, 2009; Siqueira, Boccato, Attux, & Lyra, 2014). The multilayer perceptron (MLP) is the most representative network of the FNN. Its popularity is related to the systematic training processes to tune the weights of the artificial neurons based on the backpropagation algorithm, allied with good results reported in the literature (Hippert, Pedreira, & Souza, 2001; Haykin, 2009). If the model is correctly adjusted, it can provide adequate estimations regarding a set of unknown inputs, a process known as generalization capability. Cybenko (1989) proved that an MLP, which has only one intermediate layer, is capable of approximating any nonlinear function if it is limited (the function to be approximated must be inserted in a space with boundaries to the variables) and has inputs defined in a compact space. Also, the activation functions must be infinitely differentiable (present all order derivatives needed to calculate the gradient of the cost function) (Haykin, 2009).

The unorganized machines (UM) are recent architectures of neural networks (Siqueira et al., 2014; Silva, Siqueira, Okida, Stevan Jr., & Siqueira, 2019). The difference between them and a fully trained proposal (as the MLP) is the training process. In this case, it is simple and computationally efficient as the hidden neurons stand untuned (Huang, Zhu, & Siew, 2006). Thus, the training process is reduced to find the coefficients of a linear combiner, which is the output layer. The extreme learning machines (ELM) and the echo state networks (ESN) are examples of UM.

The ELM is a single layer feedforward neural model, quite similar to the classic MLP. Be $\mathbf{u}_n = [u_n, u_{n-1}, \dots u_{n-K+1}]^T$ a vector containing the input signal. The output of a hidden layer of ELM is given by Equation 2:

$$\mathbf{x}_n^h = f^h(\mathbf{W}^h \mathbf{u}_n + b) \tag{2}$$

in which $\mathbf{W}^h \in \Re^{N_h \times K}$ is the matrix containing the weights of this layer, $f^h(\cdot) = [f_1^h(\cdot), f_2^h(\cdot), \dots f_N^h(\cdot)]$ is the activation function, and b is the bias.

The output of the network is given by Equation 3:

$$\mathbf{y} = \mathbf{W}^{out} \mathbf{x}_n^h \tag{3}$$

where:

$\mathbf{W}^{out}$ is the output weight matrix.

Due to the unorganized nature of the ELM, the training process is reduced to finding the coefficients of $\mathbf{W}^{out}$ (Huang et al., 2006). This task can be solved using the Moore-Penrose generalized inverse operation, given by Equation 4:

$$\mathbf{W}^{out} = (\mathbf{X}_h^T \mathbf{X}_h)^{-1} \mathbf{X}_h^T \mathbf{d} \tag{4}$$

where:

$\mathbf{X}_h \in \Re^{T_s \times N_k}$ as the matrix with the outputs of the intermediate layer, T_s, the number of training samples, $(\mathbf{X}_h^T \mathbf{X}_h)^{-1} \mathbf{X}_h^T$, the pseudoinverse of $\mathbf{X}_h$, and d, the vector with the desired response.

On the other hand, the ESN is a recurrent neural network (RNN) characterized by the presence of feedback loops in the intermediate layer, which is denominated dynamic reservoir (Jaeger, 2001). Following the same idea of the ELM, the weights of the reservoir are not adjusted.

Jaeger (2001) observed a few drawbacks in the training process of classic RNN, which are based in the backpropagation algorithm, such as instability, difficulties in the manipulation of the cost function, and an elevated computational cost. In this pioneer work (Jaeger, 2001), the author proves that if some conditions are satisfied, the reservoir can stand untrained and the model still presents memory, requiring only the adjustment of the output layer (Siqueira, Boccato, Luna, Attux, & Lyra, 2018; Silva et al., 2019). Additionally, Schäfer and Zimmermann (2007) proved that RNN are universal approximators, like the feedforward methods, being adequate to solve nonlinear mapping problems, such as time series forecasting.

The first design proposed for the weights of the reservoir ($\mathbf{W}^{Je}$) is randomly generating them in accordance with a predefined distribution, as given by Equation 5:

$$\mathbf{W}^{Je} = \begin{cases} 0.4 \text{ with probability } 0.025 \\ -0.4 \text{ with probability } 0.025 \\ 0.0 \text{ with probability } 0.950 \end{cases} \tag{5}$$

Finally, the adjustment process of the output layer is the same as the ELM, described in Equation 4.

Computational results and discussion

We developed two approaches considering a lag of up to 6 hours for all models (LMR, MLP, ELM, and ESN). In the first, the predictions were made using only the endogenous variables (PM_{10}, CO, and O_3). In the second, the meteorological information was considered (relative humidity - RH, average temperature - AT, and wind speed - WS) as exogenous variables, along with the endogenous ones.

During the adjustment phase of models' free parameters (training phase), we used meteorological variables and pollutant concentrations from July 17th, 2015 0 hour to March 19th, 2017 23 hour, comprising 14,688 hours (612 days). For the test set, we selected the data from March 20th, 2017 0 hour to July 17th, 2017 23 hour, containing 2,880 hours (120 days).

Performance metrics

We used the mean square error (*MSE*) and the index agreement (*IA*) to quantify the performance of the models. The *MSE* is defined in Equation 6:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y - d_i)^2 \qquad (6)$$

where:
Y is the output of the model and d_i is i-th desired response.

The index of agreement (*IA*) is a standardized measure of the degree of model prediction error and varies between 0 and 1. It is calculated according to Equation 7:

$$IA = 1 - \frac{\sum_{i=1}^{n}|Y - y_i|^2}{\sum_{i=1}^{n}(|Y - \bar{y}| + |y_i - \bar{y}|)^2} \qquad (7)$$

where:
$\bar{y}$ is the mean of all the samples of the output variable.

It is important to mention that the higher the value of *IA*, the better the model performance. That means a good prediction methodology of time series will present an *IA* close to 1. On the other hand, the *MSE* must be minimized.

Approach considering endogenous variables

To some time series, $\mathbf{u} = [u_1, u_2, \dots]$ considers the Time Series Lag Operator (TSLO) given by Equation 8:

$$L^K(u_t) = u_{t-K} \qquad (8)$$

defined for all ($t > 0$) wherein K indicates the desired lag applied on time series. That means the TSLO operates on each element of a time series to produce the Kth previous element. For example, $L^3(PM_{10})$ indicates the 3rd lag to PM_{10} pollutant concentration time series.

We obtained 63 regression equations using the PM_{10} with 6 lag hours as an independent variable and using all the possible combinations. The same method was applied to CO and O_3. The best results achieved by the LMR for PM_{10} used the respective lags $L^1(PM_{10})$, $L^2(PM_{10})$, and $L^3(PM_{10})$. For carbon monoxide, the selected lags were from the 1st to the 4th, while the ozone used $L^1(O_3)$ and $L^4(O_3)$. The neural networks presented similar configurations before the selection of the best lags. In all cases, the first lag achieved the best performance.

We also implemented the classic Autoregressive model, adjusted by the Yule-Walker Equations to provide a comparative analysis with a classical approach (Box, Jenkins, Reinsel, & Ljung, 2015). The order of the selected model followed the same as the ANN.

Moreover, we used the Mutual Information (MI) filter (Siqueira et al., 2018) to verify the nonlinear dependence between the pollutant concentration and the climate variables. In all cases, the values of the MI coefficient were higher than the threshold provided by the bootstrap method (Luna & Ballini, 2011).

Table 2 shows the computational results obtained using the cited approaches. The numbers in bold indicate the best performances for each measure. The Friedman test was applied to verify if the errors were significant. In all cases, the *p*-value was close to zero, indicating a change in the predictor leads to different results.

Comparing the *MSE* and the *IA* provided by the models, the ELM achieved the best overall results for the concentrations of PM_{10} and CO, followed by the ESN. For O_3, the MLP reached the best results, followed by LMR. This indicates that MLP and LMR should be considered to solve O_3 forecasting. It also shows that the

forecasting capability of these models is high despite the hidden layer being untuned. For CO, the performances of the UM is quite similar.

Approach considering exogenous variables

There is a specific interest in studying the influence of meteorological variables on the concentration of air pollutants since their influence by the wind speed (WS), relative humidity (RH), and temperature (AT), is noticeable (Moustris et al., 2013; Sayegh et al., 2014; Zhang & Ding, 2017). Therefore, we considered meteorological variables as input data.

Based on this, we used the IDE CodeBlocks software in view of finding the linear combination of the independent variables. The number of equations for the LMR was 15,600,380 to an exhaustive search, considering until the 6th lag hour of the current pollutant and the variables mentioned above, WS, RH, and AT. The new number of inputs was 17 for PM_{10}, 20 for CO, and 11 for O_3.

Initially, the simulations evolving the neural models, MLP, ELM, and ESN considered the 1st lag of each pollutant separately with all meteorological variables, using a total of 4 inputs. Later, we used all variables (meteorological and pollutants), presenting 6 inputs to the models. The best configuration was the same for all models: considering only 4 inputs.

We implemented the linear ARMAX model as a means of comparison (Box et al., 2015). The adjustment of the free coefficients was made using the maximum likelihood estimator. Furthermore, we used the same forward inputs and one recurrence. This is defined by previous empirical tests.

Table 3 shows the models' performance. The Friedman test ensures the statistical difference of the results.

Interestingly, all the results related to the LMR are improved by using the meteorological variables as inputs. On the other hand, the performance of the neural models was degraded. We can state that the simple inclusion of variables may not improve the performance. However, considering only the forecasting and nonlinear mapping capability of the neural models, the UM achieved, in general, best performances when compared to the MLP, except for O_3.

Figure 2, 3, and 4 allow a visual appreciation of the results, comparing the output response of the best predictors. According to them, the methods could predict the considered pollutants (PM_{10}, O_3, and CO). We highlight that the inclusion of atmospheric variables improved LMR performance. However, ANN achieved the best results without atmospheric variables.

Table 2. Prediction without meteorological variables as input data.

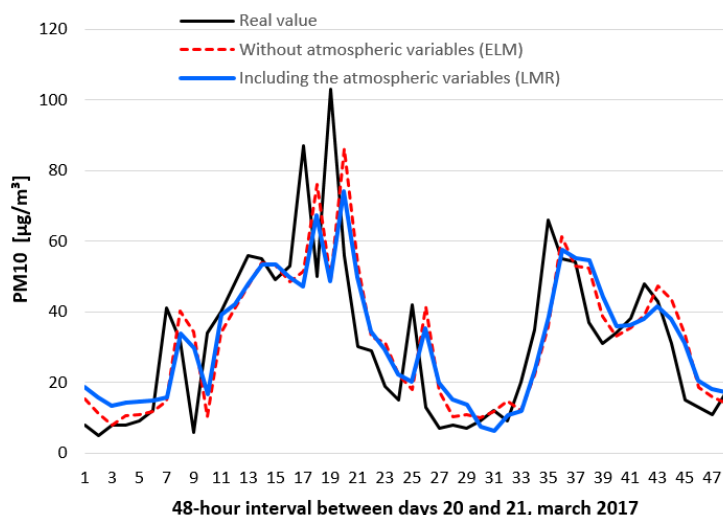|	PM_{10}		O_3		CO	
	IA	*MSE*	*IA*	*MSE*	*IA*	*MSE*
LMR	0.8721	142.7497	0.9712	8.5346	0.8992	0.0073
ELM	**0.8955**	**135.0322**	0.9699	8.7528	**0.9553**	**0.0041**
ESN	0.8943	136.9090	0.9681	9.0813	0.9547	**0.0041**
MLP	0.8770	139.9599	**0.9718**	**8.4378**	0.8769	0.0084
AR	0.8423	144.2494	0.9263	9.4129	0.8586	0.0089

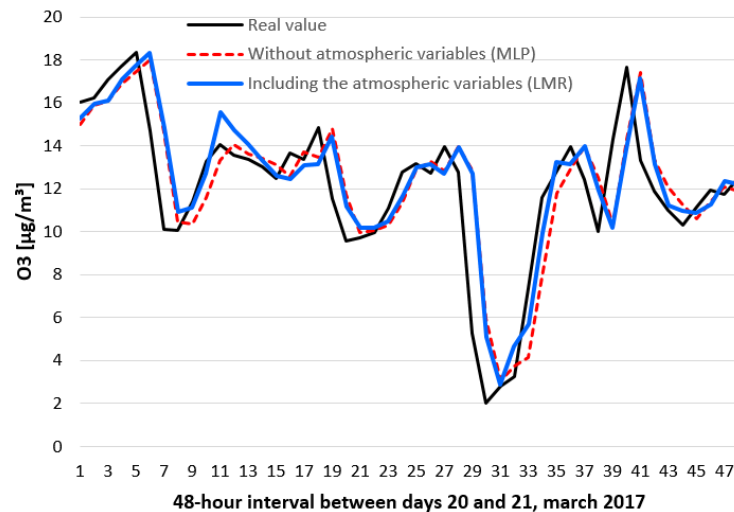**Figure 2.** Real value, ELM, and LMR predictions of PM_{10}.

Figure 3. Real value, MLP, and LMR predictions of O₃.

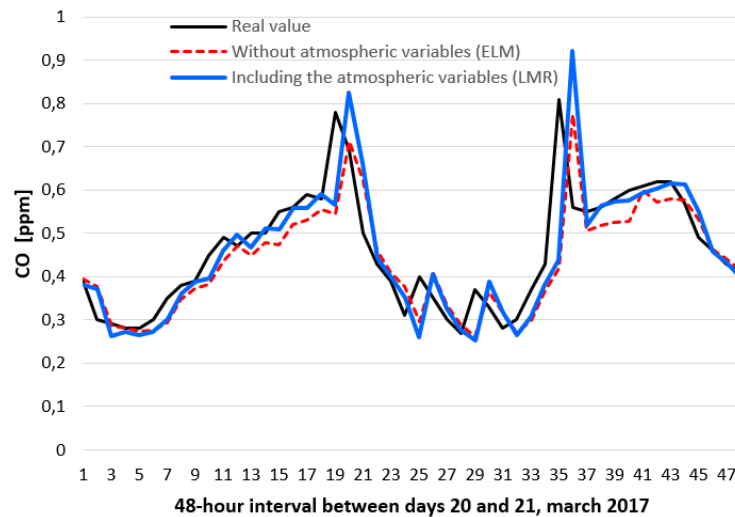**Figure 4.** Real value, ELM, and LMR predictions of CO.

Table 3. Prediction using meteorological variables as input data.

	PM_{10}		O_3		CO	
	IA	*MSE*	*IA*	*MSE*	*IA*	*MSE*
LMR	**0.8890**	**131.3932**	**0.9736**	**7.8124**	**0.9589**	**0.0044**
ELM	**0.8890**	137.5042	0.9681	9.0284	0.9413	0.0049
ESN	0.8800	146.9682	0.9618	10.4722	0.8471	0.0111
MLP	0.8643	158.6252	0.9702	8.8848	0.7894	0.0127
ARMAX	0.7997	200.3971	0.8887	15.3777	0.8328	0.0139

Conclusion

The UM showed a good performance when compared to MLP. Still, using meteorological data, the LMR presented the best results in most cases. The IA of ELM for O_3 and PM_{10} demonstrated better results. The simplicity in the training process of the UM, when compared to classic fully trained approaches provides a linear computational cost, besides the nonlinearity of the models. More accurate studies on variable selection would improve their performance. Therefore, this study should be repeated using different worldwide data sets. Finally, it is important to analyze these approaches using them as part of an ensemble or hybrid method.

Acknowledgements

References

Agência Pernambucana de Águas e Clima [APAC]. (2018). *Meteorologia*. Recuperado de http://www.apac.pe.gov.br/meteorologia/estacoes-do-ano.php?estacao=verao

Araujo, L. N., Belotti, J. T., Alves, T. A., Tadano, Y. S., & Siqueira, H. (2020). Ensemble method based on Artificial Neural Networks to estimate air pollution health risks. *Environmental Modelling & Software, 123*, 104567. doi: 10.1016/j.envsoft.2019.104567

Ardiles, L. G., Tadano, Y. S., Silvano, C., Urbina, V., Capucim, M. N., Silva, I., ..., Martins, L. D. (2018). Negative Binomial regression model for analysis of the relationship between hospitalization and air pollution. *Atmospheric Pollution Research*, 9, 333-341. doi: 10.1016/j.apr.2017.10.010

Baawain, M. S., & Al-Serihi, A. S. (2014). Systematic approach for the prediction of ground-level air pollution (around an industrial port) using an artificial neural network. *Aerosol and Air Quality Research, 14*(1), 124-134. doi: 10.4209/aaqr.2013.06.0191

Bai, Y., Zeng, B., Li, C., & Zhang, J. (2019). An ensemble long short-term memory neural network for hourly $PM_{2.5}$ concentration forecasting. *Chemosphere, 222*, 286-294. doi: 10.1016/j.chemosphere.2019.01.121

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control* (5th ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Cabaneros, S. M., Calautit, J. K., & Hughes, B. R. (2019). A review of artificial neural network models for ambient air pollution prediction. *Environmental Modelling & Software, 119*, 285-304. doi: 10.1016/j.envsoft.2019.06.014

Caselli, M., Trizio, L., Gennaro, G., & Ielpo, P. (2009). A simple feedforward neural network for the PM_{10} forecasting: comparison with a radial basis function network and a multivariate linear regression model. *Water, Air, and Soil Pollution, 201*, 365-377. doi: 10.1007/s11270-008-9950-2

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems, 2*, 303-314. doi: 10.1007/BF02551274

Dorling, S. R., Foxall, R. J., Mandic, D. P., & Cawley, G. C. (2003). Maximum likelihood cost functions for neural network models of air quality data. *Atmospheric Environment, 37*(24), 3435-3443. doi: 10.1016/S1352-2310(03)00323-6

Firmino, P. R. A., Mattos Neto, P. S. G., & Ferreira, T. A. E. (2014). Correcting and combining time series forecasters. *Neural Networks, 50*, 1-11. doi: 10.1016/j.neunet.2013.10.008

Gong, B., & Ordieres-Meré, J. (2016). Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: case study of Hong Kong. *Environmental Modelling & Software, 84*, 290-303. doi: 10.1016/j.envsoft.2016.06.020

Governo do Estado de Pernambuco. (2018). *Geografia de Pernambuco*. Recuperado de http://www.pe.gov.br/conheca/geografia/

Goyal, P., & Kumar, A. (2011). Mathematical modeling of air pollutants: an application to Indian urban city. In D. Popovic (Ed.)*, Air quality: models and applications* (p. 101-130). Rijeka, HR: InTechOpen.

Grivas, G., & Chaloulakou, A. (2006). Artificial neural network models for prediction of PM_{10} hourly concentrations, in the greater area of Athens, Greece. *Atmospheric Environment, 40*(7), 1216-1229. doi: 10.1016/j.atmosenv.2005.10.036

Haykin, S. (2009). *Neural networks and learning machines* (3rd ed.). Ontario, CA: Pearson.

Hippert, H. S., Pedreira, C. E., & Souza, R. C. (2001). Neural networks for short-term load forecasting: a review and evaluation. *IEEE Transactions on Power Systems, 16*(1), 44-55. doi: 10.1109/59.910780

Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F., & Brasseur, O. (2005). A neural network forecast for daily average PM_{10} concentrations in Belgium. *Atmospheric Environment, 39*(18), 3279-3289. doi: 10.1016/j.atmosenv.2005.01.050

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: theory and applications. *Neurocomputing, 70*(1-3), 489-501. doi: 10.1016/j.neucom.2005.12.126

Jaeger, H. (2001). *The "echo state" approach to analysing and training recurrent neural networks-with an erratum note*. Bonn, DE: German National Research Center for Information Technology GMD Technical Report.

Jasarevic, T., & Lindmeier, C. (2015). *World health assembly closes, passing resolutions on air pollution and epilepsy*. Geneva, CH: WHO. Retrieved from http://www.who.int/mediacentre/news/releases/2015/wha-26-may-2015/en/

Kachba, Y., Chiroli, D. M. G. C., Belotti, J. T., Alves, T. A., Tadano, Y. S., Siqueira, H. (2020). Artificial neural networks to estimate the influence of vehicular emissions variables on morbidity and mortality in the largest metropolis in South America. *Sustainability*, 12, 2621. doi: 10.3390/su12072621

Langrish, J. P., Li, X., Wang, S. F., Lee, M. M. V., Barnes, G. D., Miller, M. R., ... Jiang, L. X. (2012). Reducing personal exposure to particulate air pollution improves cardiovascular health in patients with coronary heart disease. *Environmental Health Perspectives, 120*(3), 367-372. doi: 10.1289/ehp.1103898

Luna, I., & Ballini, R. (2011). Top-Down strategies based on adaptive fuzzy rule-based systems for daily time series forecasting. *International Journal of Forecasting, 27*(3), 708-724. doi: 10.1016/j.ijforecast.2010.09.006

Mattos Neto, P. S. G., Madeiro, F., Ferreira, T. A. E., & Cavalcanti, G. D. C. (2014). Hybrid intelligent system for air quality forecasting using phase adjustment. *Engineering Applications of Artificial Intelligence, 32*, 185-191. doi: 10.1016/j.engappai.2014.03.010

Montgomery, D. C., & Runger, G. C. (2003). *Applied statistics and probability for engineers* (3rd ed.). New York, NY: John Wiley & Sons.

Moustris, K. P., Larissi, I. K., Nastos, P. T., Koukouletsos, K. V., & Paliatsos, A. G. (2013). Development and application of artificial neural network modeling in forecasting PM_{10} levels in a mediterranean city. *Water, Air, and Soil Pollution, 224*(8), 1634. doi: 10.1007/s11270-013-1634-x

Moustris, K. P., Nastos, P. T., Larissi, I. K., & Paliatsos, A. G. (2012). Application of multiple linear regression models and artificial neural networks on the surface ozone forecast in the greater Athens area, Greece. *Advances in Meteorology, 2012*(spe.), 894714. doi: 10.1155/2012/894714

Moustris, K. P., Ziomas, I. C., & Paliatsos, A. G. (2010). 3-day-ahead forecasting of regional pollution index for the pollutants NO_2, CO, SO_2, and O_3 using artificial neural networks in Athens, Greece. *Water, Air, and Soil Pollution, 209*, 29-43. doi: 10.1007/s11270-009-0179-5

Nagendra, S. M. S., & Khare, M. (2006). Artificial neural network approach for modelling nitrogen dioxide dispersion from vehicular exhaust emissions. *Ecological Modelling, 190*(1-2), 99-115. doi: 10.1016/j.ecolmodel.2005.01.062

Polezer, G., Tadano, Y. S., Siqueira, H. V., Godoi, A. F. L., Yamamoto, C. I., André, P. A., ... Godoi, R. H. M. (2018). Assessing the impact of $PM_{2.5}$ on respiratory disease using artificial neural networks. *Environmental Pollution, 235*, 394-403. doi: 10.1016/j.envpol.2017.12.111

Sayegh, A. S., Munir, S., & Habeebullah, T. M. (2014). Comparing the performance of statistical models for predicting PM_{10} concentrations. *Aerosol and Air Quality Research, 14*(3), 653-665. doi: 10.4209/aaqr.2013.07.0259

Schäfer, A. M., & Zimmermann, H. G. (2007). Recurrent neural networks are universal approximators. *International Journal of Neural Systems, 17*(4), 253-263. doi: 10.1142/S0129065707001111

Shahraiyni, H. T., & Sodoudi, S. (2016). Statistical modeling approaches for PM_{10} prediction in urban areas; a review of 21st-Century studies. *Atmosphere, 7*(2), 15. doi: 10.3390/atmos7020015

Silva, N., Siqueira, I., Okida, S., Stevan Jr., S. L., & Siqueira, H. (2019). Neural networks for predicting prices of sugarcane derivatives. *Sugar Tech*, *21*, 514-523. doi: 10.1007/s12355-018-0648-5

Siqueira, H. V., Boccato, L., Attux, R., & Lyra, C. (2014). Unorganized machines for seasonal streamflow series forecasting. *International Journal of Neural Systems, 24*(3), 1430009. doi: 10.1142/S0129065714300095

Siqueira, H. V., Boccato, L., Luna, I., Attux, R., & Lyra, C. (2018). Performance analysis of unorganized machines in streamflow forecasting of Brazilian plants. *Applied Soft Computing, 68*, 494-506. doi: 10.1016/j.asoc.2018.04.007

Siwek, K., & Osowski, S. (2012). Improving the accuracy of prediction of PM_{10} pollution by the wavelet transformation and an ensemble of neural predictors. *Engineering Applications of Artificial Intelligence, 25*(6), 1246-1258. doi: 10.1016/j.engappai.2011.10.013

Stadlobera, E., Hörmannb, S., & Pfeiler, B. (2008). Quality and performance of a PM10 daily forecasting model. *Atmospheric Environment, 42*(6), 1098-1109. doi: 10.1016/j.atmosenv.2007.10.073

Tadano, Y. S., Siqueira, H. V., & Alves, T. A. (2016). Unorganized machines to predict hospital admissions for respiratory diseases. *Institute of Electrical and Electronic Engineers, 2016*(LA-CCI), 1-6. doi: 10.1109/LA-CCI.2016.7885699

Tadano, Y. S., Siqueira, H. V., Alves, T. A., & Marinho, M. H. N. (2017). Forecasting particulate matter concentrations: use of unorganized machines. *International Journal of Advanced Engineering Research and Science, 4*(4), 188-191. doi: 10.22161/ijaers.4.4.27

Taylor, M., Retalis, A., & Flocas, H. A. (2016). Particulate matter estimation from photochemistry: a modelling approach using neural networks and synoptic clustering. *Aerosol and Air Quality Research, 16*(9), 2067-2084. doi: 10.4209/aaqr.2015.07.0481

Wang, J., & Song, G. (2018). A deep spatial-temporal ensemble model for air quality prediction. *Neurocomputing, 314*, 198-206. doi: 10.1016/j.neucom.2018.06.049

World Health Organization [WHO]. (2006). *Air quality guidelines. Global update 2005. Particulate matter, ozone, nitrogen dioxide and sulfur dioxide*. Copenhagen, DK: WHO.

Wu, C. F., Li, Y. R., Kuo, I. C., Hsu, S. C., Lin, L. Y., & Su, T. C. (2012). Investigating the association of cardiovascular effects with personal exposure to particle components and sources. *Science of the Total Environment, 431*, 176-182. doi: 10.1016/j.scitotenv.2012.05.015

Zhang, J., & Ding, W. (2017). Prediction of air pollutants concentration based on an extreme learning machine: the case of Hong Kong. *International Journal of Environment Research and Public Health, 14*(2), 114. doi: 10.3390/ijerph14020114