Environmental Modelling and Software 123 (2020) 104567

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: http://www.elsevier.com/locate/envsoft

Ensemble method based on Artificial Neural Networks to estimate air pollution health risks

Lilian N. Araujo [a,b], Jônatas T. Belotti [b], Thiago Antonini Alves [b], Yara de Souza Tadano [b], Hugo Siqueira [b,*]

[a] Federal Institute of Parana (IFPR), Brazil
[b] Federal University of Technology - Parana (UTFPR), Brazil

ARTICLE INFO

Keywords:
Artificial neural networks
Ensemble method
Generalized linear model
Particle swarm optimization
Hospital admissions
PM_{10}

ABSTRACT

Estimating of daily hospital admissions due to air pollution is a leading issue in environmental science. To better understand this problem, it is essential to improve the applied methodologies. The use of Generalized Linear Models (GLM) is well known. However, they may be improved using different methods to coefficients estimation and to consider seasonality. Alternative methodologies, rarely applied in such topic, are Artificial Neural Networks (ANN), efficient to solve non-linear problems and; ensembles, which combine various models outputs. This research aims to apply 10 distinct ANN and 4 ensemble to estimate hospital admissions for respiratory diseases caused by particulate matter and meteorological variables of Campinas and São Paulo cities, Brazil. In addition, a new proposal of GLM was introduced, considering coefficients calculation via particle swarm optimization and seasonality via normalization procedure. ANN and ensembles use showed significant improvements and may allow studies into areas with flawing database, developing countries reality.

1. Introduction

The air pollution impact on human health is a topic discussed worldwide. The World Health Organization (WHO) reports that more than 6 million premature deaths are related to the pollution (WHO, 2017). To better understand the association between air pollution and adverse health effects, it is essential to improve the applied methodologies.

Air pollution impact on human health is commonly performed using statistical regression techniques, especially the Generalized Linear Models (GLM) (Tadano et al., 2012) (Vanos et al., 2014) (Polezer et al., 2018) (Pan et al., 2019) (Kim and Lee, 2019). GLM usually use maximum likelihood estimators of regression coefficients (Dobson and Barnett, 2008a) (McCullagh and Nelder, 1989), the default option of statistical softwares such as Splus, SAS, Stata and R.

An alternative for regression coefficients estimation may consist on, instead of maximum likelihood estimators, bio-inspired metaheuristics as the Particle Swarm Optimization (PSO), which presents a search capability that may achieve better accuracy in the coefficients tuning than traditional methods (Matsuzaki, 2017).

Another improvement may be achieved for seasonality trends, switching the use of smoothing splines with monthly deseasonalization by normalization technique, a simpler method with potential of transforming the variables into stationary (Matsuzaki, 2017) (Siqueira et al., 2014) (Siqueira et al., 2018).

The use of statistical regression techniques to assess adverse health effects of air pollution requires at least two years of consistent and robust databases (Tadano et al., 2012). A minimum of missing data is needed, otherwise, the model may not capture a relationship between the inputs (independent variables) and the output (dependent variable), due to seasonal tendencies of the series (Tadano et al., 2012) (Polezer et al., 2018), turning this kind of studies a true challenge in developing countries, that usually have incomplete databases.

An alternative methodology consists on Artificial Neural Networks (ANN), widely used in many different types of predictive modeling, including air pollution mapping ((Mishra et al., 2015) (Franceschi et al., 2018) (Feng et al., 2019) (Cabaneros et al., 2019)). Cabaneros, Calautit and Hughes (Cabaneros et al., 2019) reported a review about researches aiming to forecast air pollution using ANN from 2001 to February 2019, resulting in 139 peer-reviewed articles. However, the application of

* Corresponding author.
E-mail addresses: lika.n.araujo@gmail.com (L.N. Araujo), jonatas.T.Belotti@hotmail.com (J.T. Belotti), thiagoaalves@utfpr.edu.br (T.A. Alves), yaratadano@utfpr.edu.br (Y.S. Tadano), hugosiqueira@utfpr.edu.br (H. Siqueira).

https://doi.org/10.1016/j.envsoft.2019.104567
Received 14 April 2019; Received in revised form 11 October 2019; Accepted 17 October 2019
Available online 23 October 2019

ANN to estimate potential air pollution impacts on population health is less used ((Polezer et al., 2018) (Wang et al., 2008) (Kassomenos et al., 2011) (Tadano et al., 2016) (Sundaram et al., 2016)). Also, the frequently used ANN consists of Multilayer Perceptron (MLP).

Feedforward ANN, as the MLP, are nonlinear methodologies that are capable of approximate any nonlinear continuous function, if it is limited, differentiable and, present inputs defined in a compact space with arbitrary precision. In other words, it means that this kind of method can map a set of inputs in an output, using a nonlinear way, being known as universal approximators, which present elevated generalization capability. This is exactly the case of the problem we addressed. In addition, Recurrent ANN are general cases of the Feedforward ones. In essence, ensembles are extensions of the ANN because they only combine the outputs of the previous methods.

In epidemiological studies, the Generalized Linear Models frequently do not achieve satisfactory performance due to the inherent process of free coefficients adjustment (Coman et al., 2008) (Lauret et al., 2016) (Chelani et al., 2002).

In recent times, the possibility of using ensembles has gained much attention due to their accuracy and efficiency, because they combine the output of **multiple single models** responses. Trainable ensembles are those in which the combiner is, for example, an adjustable model like a neural network, while a non-trainable method consists of a statistical measure of the inputs, as the average (de Mattos Neto et al., 2014) (Firmino et al., 1016).

With this in mind, this research proposes the application of ANN, neural-based ensembles and the aforementioned new version of the GLM to estimate the impact of air quality on population health. The addressed ANN were: multilayer perceptron (MLP), Radial Basis Function Network (RBF), Extreme Learning Machines (ELM), Echo State Networks (ESN), Jordan and Elman networks. The goal was to verify the contribution of these approaches to estimate the number of hospital admissions due to air pollution exposure. The case studies consider particulate matter with aerodynamic diameter less than or equal to 10μm (PM_{10}) and meteorological variables influence on hospital admissions for respiratory diseases at two important Brazilian cities: Campinas and São Paulo cities.

2. Generalized linear model

Generalized Linear Models (GLM) are an extension of the classical linear model used for continuous problems. This method describes the relationship between one or more prediction variables (independent, explanatory or covariable) ($\mathbf{x} = x_1, \ldots, x_p$), for a wide variety of non-distributed responses (Neuhaus and McCulloch, 2011). A key point on GLM use is that many of the considerations in its construction are the same for the standard linear regression models, since both have many characteristics in common (Matsuzaki, 2017).

The GLM presents three specifications: the random component, or the distribution of the response; the systematic component - the function of the prediction variables; and the connection between the other components (Neuhaus and McCulloch, 2011). Therefore, GLM assumes that the response refers to the expected value of the target variable, achievable through a linear predictor. It is supposed that the model manipulates a combination of inputs and can use polynomial functions or splines (which is a smoothing that approximates functions behavior that have local and sudden changes) (Tadano et al., 2012).

The Poisson regression is a specific approach of the GLM, which further explains that the response variable must follow the same distribution, and the data must have equal dispersion (Neuhaus and McCulloch, 2011). According to (Matsuzaki, 2017), in the Poisson regression model, the response variable describes the number of times that an event occurs in a finite observation space. Also, the function given by Equation (1) is the inverse of the function (log α) called the link function:

$$\alpha = e^{\beta_0+\beta_1 x_1+\ldots+\beta_n x_n} \tag{1}$$

in which α is the predicted value, β_0 plays the role of the offset and $\beta_1, \ldots, \beta_n$ are the free coefficients that are multiplied by the inputs vector ($\mathbf{x} = x_1, \ldots x_n$).

2.1. New GLM

The free coefficients of GLM with Poisson regression are commonly estimated using the maximum likelihood method (Paula, 2004) (Dobson and Barnett, 2008b). However, it is possible to tune the model using bio-inspired metaheuristics, like the Particle Swarm Optimization (PSO).

The PSO was inspired in the intelligent collective behavior of animals groups, such as school of fishes and flocks of birds. This technique was introduced by Eberhart and Kennedy (1995). The method works with a group of agents moving through the search space, trying to identify the optimal points of the function to be optimized. Each agent is an independent particle, or a candidate solution, and during the search process it works together with other agents, forming a population. In this work, each particle is composed by the GLM free parameters.

During the displacement, the particles move towards the best position they have reached (**pbest**) and in the direction of the best position achieved by some of its neighbors (**gbest**).

The particle $i = 1, \ldots, P$ has a position vector x with $d = 1, \ldots, D$ dimensions, which are changed by the speed operator v. These values are updated as shown in Equations (2) and (3):

$$v_{id}^{k+1} = \omega v_{id}^{k} + c_1 rand()\left(pbest_{id} - x_{id}^{(k)}\right) + c_2 rand()\left(gbest_{d} - x_{id}^{(k)}\right) \tag{2}$$

$$x_{id}^{k+1} = x_{id}^{(k)} + v_{id}^{(k+1)} \tag{3}$$

in which, $v_{id}^{(k)}$ is the velocity of the individual in the iteration k, $v_d^{min} \leq v_{id}^{(k)} \leq v_d^{max}$, ω is the inertia weight factor, c_1 e c_2 are the acceleration constants, *rand*(): uniformly generated random numbers between 0 and 1, and $x_{id}^{(k)}$ is the current position of the individual in iteration k.

Besides the GLM tuning, another important remark is to consider the data seasonality. According to Tadano et al. (Tadano et al., 2012), it is an important long-term trend that appears in this kind of study, as the meteorological variables and concentration of pollutants vary during the year. To adjust these components, the spline function is often used, because it provides a better approximation than simple polynomial interpolations (Markatos, 1985).

However, it is possible to replace the spline procedure by the deseasonalization method (Siqueira et al., 2014) (Siqueira et al., 2018). It makes the series stationary, by means of the application of Equation (4) (Siqueira and Luna, 2019)

$$z_i = \frac{x_i - \mu_m}{\sigma_m} \tag{4}$$

being μ_m the mean and σ_m the standard deviation of each month.

The new series z_i is approximately stationary, with zero mean and unit standard deviation.

3. Artificial neural networks

In this study we addressed feedforward networks (multilayer perceptron - MLP and radial basis function networks - RBF), recurrent architectures (Elman and Jordan networks), and the unorganized machines (Extreme Learning Machines and the Echo State Networks) (Siqueira et al., 2014) (Polezer et al., 2018) (Haykin et al., 2009).

3.1. Multilayer perceptron

The multilayer perceptron (MLP) is considered one of the most important architectures of ANN. This method presents a set of

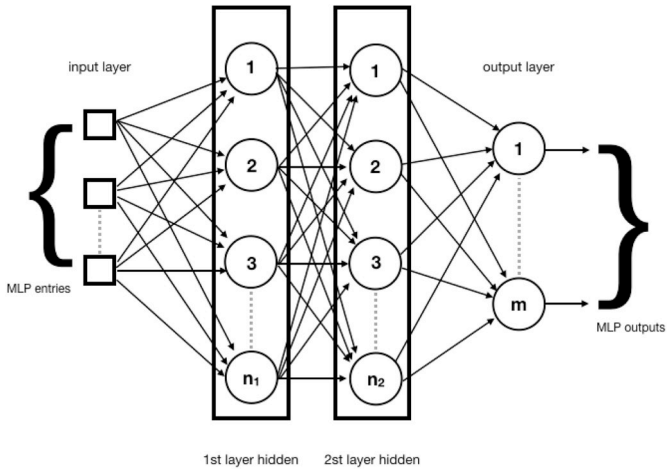

Fig. 1. MLP structure.

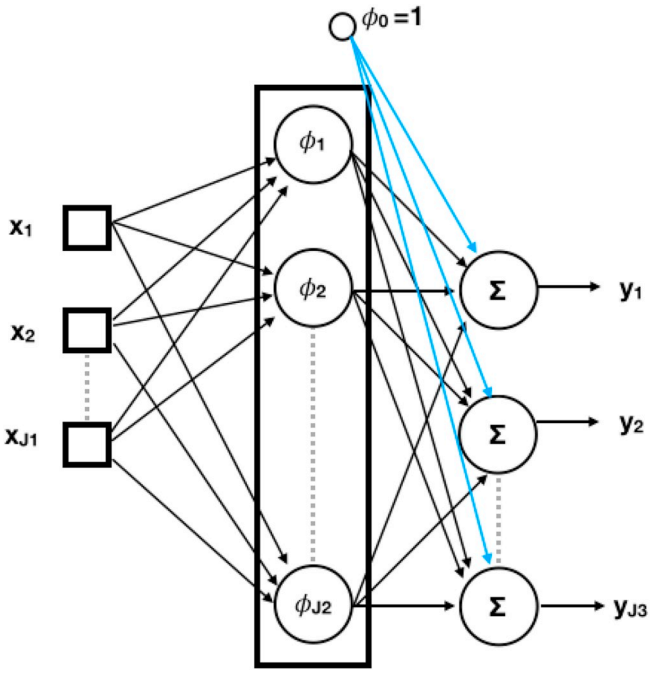

Fig. 2. RBF network.

information processors units known as artificial neurons, which form the input layer, one or more hidden layers, and an output layer, as shown in Fig. 1 (Haykin et al., 2009).

The information flows from the input layer, passes through the intermediate layers, and closes in the output layer, generating an output response. In a MLP the disjoint layers are connected, whereas neurons of the same layer do not communicate. The training, or the adjustment of their weights, is performed in two phases: the first one is the forward propagation, in which the signals from a training set sample are inserted as inputs of the network, being propagated layer by layer; in the second phase, the errors are propagated in a recursive manner, and the weights are adjusted through some adjustment rule (Haykin et al., 2009) (Siqueira et al., 2014). The most commonly used method to tune the MLP is the steepest descent algorithm, with the derivatives calculated via the backpropagation method.

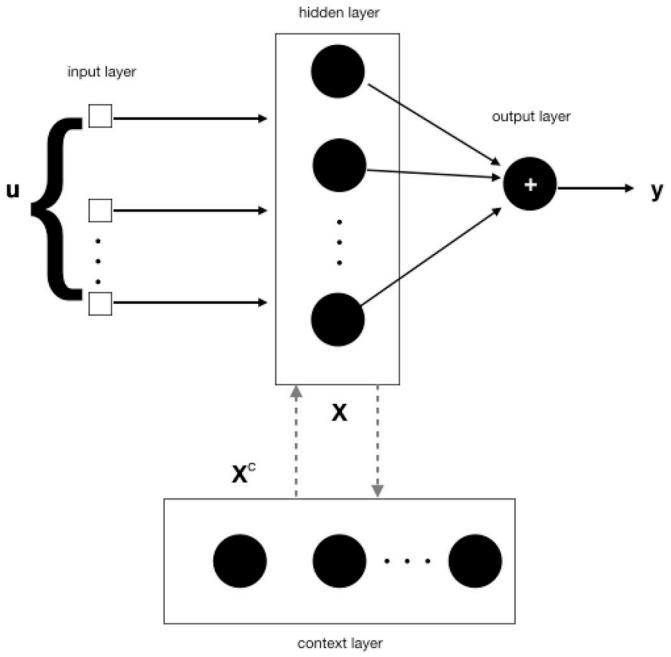

Fig. 3. Elman network structure.

3.2. Radial basis function networks

Radial Basis Function Networks (RBF) present their origin in the execution of exact interpolation of a set of data points in a multidimensional space (Powell, 1987). These ANN have a simpler structure and a faster training process compared to MLP. It is composed of only one intermediate layer, in which the neurons present activation functions different from the output layer neurons. In this case, the activation function inputs are the Euclidean distances between the input information and the weight vectors. Fig. 2 shows the RBF architecture.

In a RBF, each neuron in the hidden layer uses a nonlinear radial basis function as activation function. This layer performs a nonlinear transformation of the input. In this sense, the output layer acts as a linear combiner that maps the nonlinearity to a new space. The bias of the output layer neurons can be modeled by an additional signal in the previous one, which has a constant activation function (Du and Swamy, 2014).

The RBF training consists of two steps. The first one is associated to the adjustment of the weights of the intermediate layer neurons and uses unsupervised learning methods, directly related to radial basis functions. The second step is the adjustment of the output layer weights which uses, for example, the generalized delta rule or the backpropagation method, similar to the MLP.

3.3. Neural network of elman and Jordan

Recurrent Neural Networks (RNN) are those endowed with feedback loops of information. In these cases some output of hidden or output layers are reinserted in some previous one. This process provides to the network a kind of internal memory, which can be useful to problems with time dependency (Haykin et al., 2009).

The Elman neural network (ENN) is a recurrent model, which presents additional inputs in the hidden layer forming the context layer. Fig. 3 describes the Elman network with four layers: input, hidden (intermediate), context and output layers (Ren et al., 2018). The context layer stores the output values of the hidden layer of the previous time. Also, excepting the context layer, the model is similar to a feedforward network.

In Fig. 3, the inputs are represented by $\mathbf{u}_k$, the outputs are $\mathbf{y}_k$, the

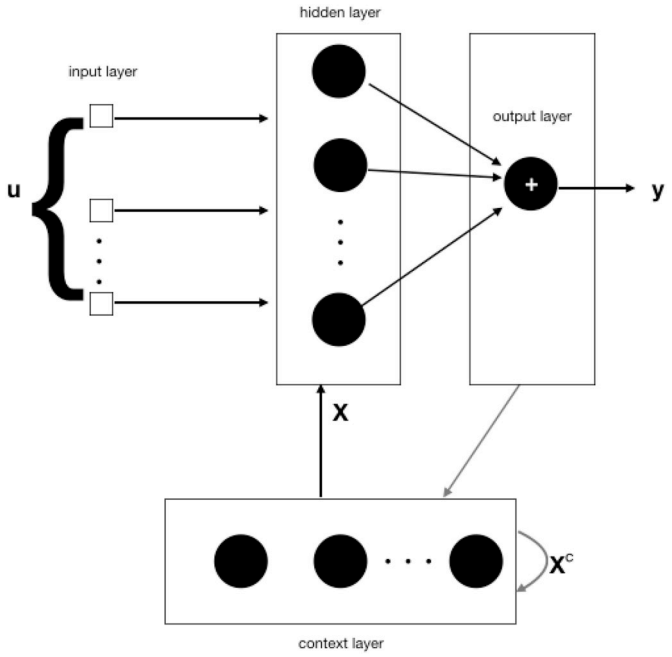

Fig. 4. Jordan network structure.

entries for i-th neuron of the hidden layer are denoted as v_k^i, the output of the i-th hidden layer is given by x_k^i, and the output of the j-th context is $\mathbf{x}_k^{c,j}$.

The standard method to training an ENN is called the Elman back-propagation algorithm (EBP) (Ren et al., 2018), in which the time is treated as an explicit part of the input. It allows time to be represented by the effect it has on the process, which gives dynamic properties to the system, which respond to temporal sequences (Elman, 1990). In short, the network becomes endowed with memory. When the back-propagation is applied, the dependence of $\mathbf{x}^c(k)$ to the weights must be taken into account. The algorithm that deals with this dependency is the dynamic backpropagation algorithm (Kuan, 1989).

Local RNN stores the past state of the networks and use it as part of the entry in the next iteration. They can be divided into two categories: a network with external Time Delay Feedback and a network with Internal Delay Feedback. The first case is called Jordan neural network (JNN) (Ren et al., 2018).

The main difference, in terms of the structure between the ENN and JNN is shown in Fig. 4. While the in an ENN the recurrent information comes from the intermediate layer to the input, in the JNN the recurrence is from the output layer to the input (Haykin et al., 2009).

3.4. Unorganized machines

Unorganized machines (UM) are architectures of artificial neural networks, in which the weights of the intermediate layer are not adjusted, and are randomly generated. Because of this, their training process are simple and computationally efficient. The term Unorganized Machines was proposed by Boccato et al. (2011), which evokes Alan Turing's work on intelligent machine behavior (Turing). In this work, we addressed the Extreme Learning Machines and the Echo State Networks.

3.4.1. Extreme Learning Machines

According to Huang (Huang et al., 2006), the Extreme Learning Machines (ELM) are similar to MLP networks, with a single intermediate layer. However, in this case, the weights of the neurons are chosen in a random and independent way. The adjustment process is performed only on the output layer, which is, a linear combiner. Therefore, the

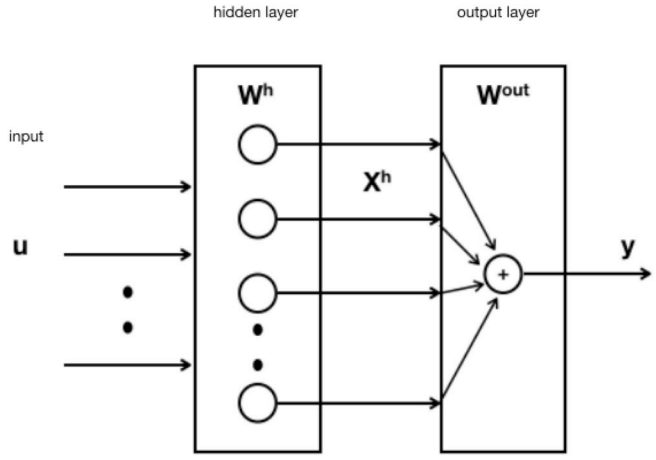

Fig. 5. ELM structure.

training process has a linear computational cost and uses a supervised approach. Fig. 5 shows the structure of a generic ELM with its layers. Here, it is represented only one output.

The input vector is $\mathrm{u}_t = [u_1, u_2, \ldots, u_{t-K+1}]^T$. Its components passed to the hidden layer $\mathrm{W}^h \in \Re^{N_h \times K}$, which contains the random generated weights. The output signal of this layer is according to Equation (5):

$$\mathbf{x}_t^h = \mathbf{f}^h\left(\mathbf{W}^h \mathbf{u}_t + \mathbf{b}\right) \tag{5}$$

in which $\mathbf{b}$ is the bias of each neuron and $\mathbf{f}^h(.)$ is the activation function of the intermediate neurons. Such activations are linearly combined to produce the network output y_t, using Equation (6):

$$y_t = \mathbf{W}^{out} \mathbf{x}_t^h, \tag{6}$$

being $\mathbf{W}^{out}$ the matrix of the output layer weights.

The training process consists of finding the best weights in the output layer. The Moore-Penrose operator is an interesting option to solve the task:

$$\mathbf{W}^{out} = \left(\mathbf{X}_h^T \mathbf{X}_h\right)^{-1} \mathbf{X}_h^T \mathbf{d}, \tag{7}$$

where $\mathbf{X}_h \varepsilon \Re^{T_s \times N_k}$ is the matrix containing the exit of the intermediate layer, T_s is the number of samples of the training, $\left(\mathbf{X}_h^T \mathbf{X}_h\right)^{-1} \mathbf{X}_h^T$ is a pseudo-inverse of $\mathbf{X_h}$ and $\mathbf{d}$ is the vector containing the desired output.

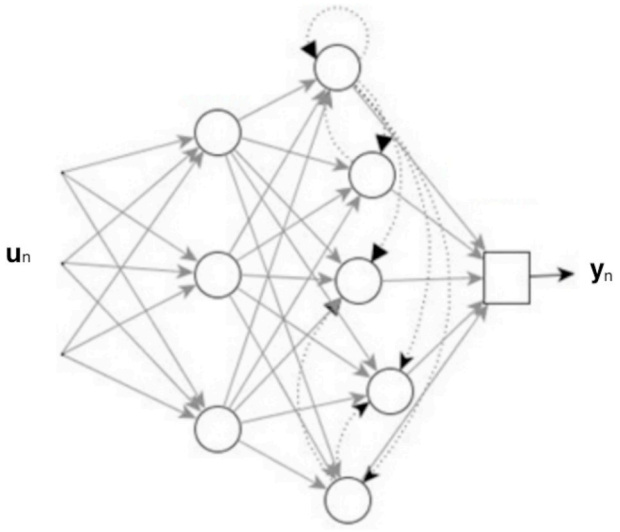

Fig. 6. Echo state network.

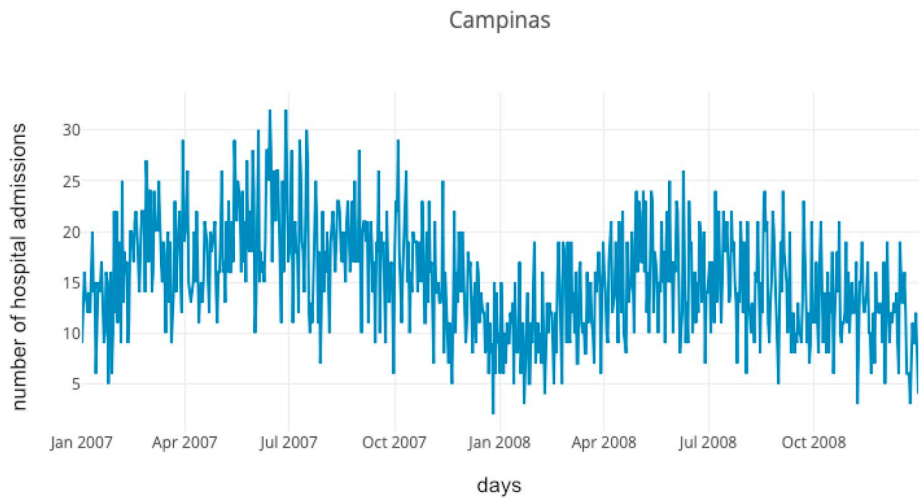

Fig. 7. Number of hospitalizations per period - Campinas.

In (Huang et al., 2006), the authors suggest that the generalization capability of the network can be increased by inserting a weighting term C to Equation (7), as in 8.

$$\mathbf{W}^{out} = \left(\frac{I}{C} + \mathbf{X}_h^T \mathbf{X}_h\right)^{-1} \mathbf{X}_h^T \mathbf{d}, \tag{8}$$

To determine the value of C, the authors indicate to test the 52 possible values of the vector $\lambda = \{-26, -25, \ldots, 24, 25\}$, so that $C = 2^{\lambda}$. Thus, after training, a validation set must be used. To this term C, the authors gave the name Regularization Coefficient.

3.4.2. Echo State Networks

Echo State Networks (ESN) were proposed by Jaeger (2001). They are recurrent structures because they have feedback loops between the neurons, which differs it from the ELM. This feature allows the creation of an intrinsic memory, which may be beneficial to problems involving temporal relation between samples.

The training process is similar to that of the ELM, which must find the coefficients of a linear combiner based on a least squares problem with reference signal. Fig. 6 shows the generic structure of an ESN, in which the input vector is $\mathbf{u}_n = [u_n, u_{n-1}, \ldots, u_{n-k+1}]^T$, which is transmitted from the input layer $\mathbf{W}^{in}$ to the dynamic reservoir (intermediate layer) by means of non-linear combinations.

The theoretical element that guarantees the presence of memory in the ESN is known as the echo state property (Jaeger, 2001). Under specific conditions, in the reservoir $\mathbf{W}$, the recent history of the inputs governs the internal dynamics of this layer neurons. The network weights can thus be previously defined (Jaeger, 2001; Yildiz et al., 2012).

Compared to the other RNN, the training process of ESN is simpler and faster. There are practically no problems related to the instability and manipulation of cost functions (Siqueira et al., 2014). As in the ELM, the task is to find the coefficients of a linear combiner using the Moore-Penrose inverse operator and the Regularization Coefficient.

We highlight that the design of the reservoir plays an important role in the development of the ESN. In this work we addressed the proposals from (Jaeger, 2001) and from Ozturk et al. (2007), following the steps described in (Siqueira et al., 2014).

4. Ensembles

The literature has shown that the use of ensembles brings performance enhancements to the single models (de Mattos Neto et al., 2014; Firmino et al., 1016; Perrone and Cooper, 1992). According to (Perrone and Cooper, 1992), ensembles combine the outputs of various models, which were pretuned. In the beginning, the application of this technique was given by the weighted average calculation of each methods outputs. The approach is known as Basic Ensemble Method (BEM) and is given by 9:

$$f_{BEM} = \frac{1}{n}\sum_{i=1}^{n} f_i \tag{9}$$

where n is the total number of single models and f_i is the output of these models.

However, in recent times, several works have presented the possibility of using a neural network as the combinator (de Mattos Neto et al., 2014) (Firmino et al., 1016). The idea can be exemplified as follows: consider that there are three different ANN to estimate a given variable. The neural-based ensemble will use the point-to-point predictions of each network and will use them as input. Then, the combinatorial network must be trained in the same traditional way. This process tends to lead to better solutions. In this work, we will use the MLP and RBF networks to perform this task because they are the most important feedforward ANN models, and other studies addressed them in correlated problems (de Mattos Neto et al., 2014) (Firmino et al., 1016).

5. Case study

The case study consists of estimating the number of hospital admissions due to respiratory diseases caused by the exposure to particulate matter (PM_{10}) to Campinas and São Paulo cities, Brazil. According to data from the Brazilian Institute of Geography and Statistics (IBGE), Campinas had 1,194,094 inhabitants in 2017, being the third most populous municipality in the state of São Paulo. It occupies an area of 797.6 km^2, of which 238.3 km^2 are in the urban perimeter and the remaining 559.3 km^2 is the rural area (de Campinas, 2014). It presents a tropical climate of high altitude, with hot and rainy summers. The temperatures are in the range from 13 °C to 29 °C, being rarely less than 9 °C or greater than 33 °C (Spark, 2018).

São Paulo city is the capital of the state of São Paulo and the main business and commercial financial center of South America. It is the most populous city in Brazil with 12,176,866 inhabitants, according to IBGE (IBGE, 2019). It is also the most influential in the global scenario, being considered in 2016, the 11th most globalized city on the planet. It is 464 years old, and has a total area of 1521.11 km^2, being 968.32 km^2 of urban areas.

The information about hospital admissions for respiratory diseases were obtained from the Brazilian National Health System at http://www2.datasus.gov.br/DATASUS/index.php?area=02 (DATASUS, 2010).

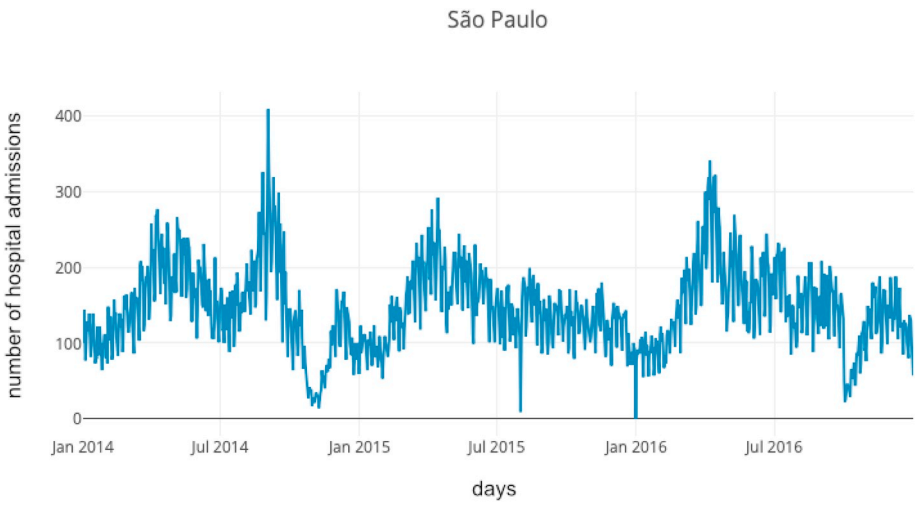

Fig. 8. Number of hospitalizations per period - São Paulo.

The database from Campinas used for the models adjustment and estimation has a total of 732 samples, comprising the period from January 1st, 2007 to December 31st, 2008. During that period, 11,121 hospitalizations were detected, being 2 the smaller number and 32 the greater to a single day, as shown in Fig. 7. São Paulo dataset has 1068 samples, in which the number of hospitalizations lies from January 1st, 2014 to December 31st, 2016 totaling 159,683 occurrences. The highest daily number of hospitalizations was 409, as can be observed in Fig. 8. On the other hand, there were days with no attendance regarding respiratory diseases, probably due to a matter of database record system or the fact the available data is only from public health system, miss considering data from health insurances and private units.

The pollution concentration (PM_{10}) and meteorological data (ambient temperature and relative humidity) are from air quality monitoring stations administered by Environmental Company of São Paulo State (CETESB) at https://cetesb.sp.gov.br/ar/qualar/(CETESB, 2010). Due to lack of resources, São Paulo government had only one air quality monitoring station (AQMS) at Campinas city. Nowadays, they have one more station that has PM_{10} data, but only since 2016 (CETESB, 2018). PM_{10} and meteorological dataset consists of daily averages from the Cerqueira Cesar station, held by CETESB. To the studied period only four out of twelve stations had PM_{10} available data. Even more, only one station has less than 100 days lack. Due to the ligh linear correlation between data from these four stations, we chose data from Cerqueira Cesar Station, with only 29 days lack.

Besides that, this study considered as an independent input the day of the week and whether the day of hospitalization is a holiday or not. These two variables were used because they follow a pattern in epidemiological studies (Tadano et al., 2016), since the number of hospitalizations decreases on weekends and holidays (Polezer et al., 2018). In this way, the entries of the models covered, contain the following fields:

- Date - Date referring to the day of measurement;
- PM_{10} - Concentration of particulate matter with aerodynamic diameter less than or equal to 10μm;
- Temperature - Recorded Average temperature;
- Humidity - Recorded Air Relative humidity;
- Day - Day of the week (from Sunday -1- to Saturday -7-);
- Holiday - Whether the day is a holiday -1- or not -0-;
- Number of hospitalizations - Number of hospital admissions due to respiratory problems according to the International Classification of Diseases (ICD-10 - J00 to J99).

In both cases, the samples were divided into three sets: training containing the first 70% of the data, validation, with the following 15% and the test set, with the remaining 15%. That percentage was tested being reasonable by the amount of available data.

Table 1
Parameters of the PSO applied to the GLM.

Parameters	Value
Number of particles	20
Inertia weight, ω	0.5
Cognitive parameter, c_1	2
Social parameter, c_2	2

5.1. Results and analyzes

To perform an extensive and significant analysis of the models mapping capability, the following neural network architectures were elaborated: MLP, RBF, Elman network, Jordan network, ELM, ESN - Jaeger, ESN - Ozturk. Also, we perform the models that have the regularization coefficient in their structure: ELM (RC), ESN-Jaeger (RC), ESN-Ozturk (RC). In addition, the ensembles were implemented considering the linear average, median, MLP and RBF as combinators. Finally, we used the traditional GLM model (with coefficients calculated by the maximum likelihood estimator and long-term trend smoothing with splines) and the new proposal optimized by the PSO, and using the deseasonalization procedure. The parameters of the PSO were chosen based on previous empirical tests, and they are in Table 1.

All models were developed using the Java programming language with the object-oriented paradigm under Netbeans IDE. The exception was the traditional GLM, which was performed using the R ®software. In this case, in order to adjust seasonal and long-term trends, we used the spline function toolbox from the R, with value equals to 16 to both cities, value based on previous studies in this field.

The ANN were elaborated using only one intermediate layer. In the training, the amount of artificial neurons was selected in the range from 5 to 200, with increments of 5. The weight adjustment was performed in the training set. The cross-validation procedure was applied in view to increase the generalization capability of the networks (Haykin et al., 2009), except to the ELM and ESN without RC. The training stop condition was a precision of 10^{-6} with a maximum of 2000 iterations.

It is important to mention that all neural networks had the linear identity function in the output neurons as activation function, and the hyperbolic tangent in the hidden layer, except RBF. In the last case, the K-Means algorithm was applied to cluster the centers. In addition, the weights were generated randomly in the interval [-1; 1], as well as the

Table 2
Results for campinas.

Lag	0				1			
Model	NN	MSE	MAE	MAPE	NN	MSE	MAE	MAPE
MLP	20	12.212	2.888	27.47	10	11.658	2.764	26.36
RBF	10	12.620	2.864	30.52	30	11.288	2.715	28.67
ELM	20	13.729	3.036	32.23	30	13.878	3.039	33.58
ELM CR	50	14.818	3.130	33.95	35	14.404	3.083	32.91
Elman	15	16.575	3.339	32.55	5	12.991	2.913	27.64
Jordan	5	12.362	2.811	26.74	10	11.412	2.728	26.37
ESN Ozturk	45	13.884	3.070	32.69	25	13.693	3.076	33.47
ESN Jaeger	25	13.793	3.050	32.98	80	13.260	2.925	31.14
ESN Ozturk CR	15	14.874	3.188	35.18	50	14.127	3.070	32.61
ESN Jaeger CR	15	14.431	3.152	33.66	40	14.318	3.058	33.63
GLM	–	21.487	3.873	39.23	–	23.845	3.998	29.49
GLM/PSO	–	19.406	3.629	42.04	–	19.986	3.656	42.58
Ensemble 1	–	12.258	2.896	30.95	–	12.228	2.850	31.06
Ensemble 2	–	11.978	2.859	30.12	–	11.853	2.806	30.19
Ensemble 3	5	11.991	2.737	26.32	**105**	**10.637**	**2.671**	**26.48**
Ensemble 4	**185**	**11.069**	**2.758**	**27.56**	25	11.306	2.719	28.33
Lag	2			3				
MLP	5	11.733	2.746	27.07	5	11.815	2.743	27.26
RBF	10	11.788	2.759	29.50	145	11.150	2.747	29.02
ELM	65	12.175	2.812	29.43	70	12.855	2.883	30.26
ELM CR	65	13.171	2.971	31.52	45	13.341	2.990	31.28
Elman	5	12.965	2.904	27.42	5	13.308	2.914	27.23
Jordan	5	11.654	2.729	25.79	5	12.002	2.743	27.41
ESN Ozturk	85	12.461	2.799	29.47	50	13.145	2.932	30.52
ESN Jaeger	65	12.784	2.910	31.54	30	13.484	3.007	33.24
ESN Ozturk CR	80	12.811	2.843	29.15	15	14.951	3.205	34.95
ESN Jaeger CR	55	13.694	3.020	31.88	70	13.579	2.991	31.84
GLM	–	19.695	3.641	28.20	–	37.623	5.092	35.89
GLM/PSO	–	19.510	3.614	42.19	–	19.808	3.659	42.46
Ensemble 1	–	11.721	2.782	30.03	–	11.678	2.802	29.91
Ensemble 2	–	11.358	2.733	29.05	–	11.342	2.762	29.01
Ensemble 3	**25**	**10.291**	**2.535**	**25.36**	**175**	**9.298**	**2.439**	**24.87**
Ensemble 4	15	11.298	2.675	28.51	25	11.197	2.701	27.33
Lag	4			5				
MLP	5	12.679	2.834	26.93	25	11.314	2.732	26.67
RBF	30	12.095	2.818	30.08	15	11.993	2.809	29.90
ELM	60	12.986	2.938	30.76	60	12.545	2.839	29.31
ELM CR	40	13.940	3.068	31.94	70	12.772	2.912	30.87
Elman	5	13.757	2.966	27.65	5	13.056	2.930	28.05
Jordan	5	12.270	2.723	25.99	20	11.943	2.765	26.39
ESN Ozturk	85	12.591	2.865	29.39	60	12.026	2.847	30.41
ESN Jaeger	85	12.773	2.888	29.78	25	12.593	2.883	30.13
ESN Ozturk CR	60	13.664	3.040	31.79	20	14.324	3.104	32.93
ESN Jaeger CR	55	14.365	3.071	31.95	95	12.010	2.792	28.93
GLM	–	19.829	3.664	28.28	–	20.549	3.721	28.26
GLM/PSO	–	18.302	3.529	40.54	–	17.095	3.459	39.28
Ensemble 1	–	12.167	2.844	30.37	–	11.288	2.742	29.11
Ensemble 2	–	11.880	2.795	29.37	–	10.935	2.685	28.01
Ensemble 3	**5**	**11.247**	**2.623**	**27.15**	**10**	**10.220**	**2.527**	**25.66**
Ensemble 4	10	11.856	2.838	29.40	35	10.964	2.659	29.41
Lag	6			7				
MLP	15	12.042	2.761	27.78	5	11.647	2.690	26.12
RBF	15	11.969	2.819	29.46	15	11.582	2.779	29.54
ELM	40	13.263	3.001	32.35	85	12.997	2.983	31.58
ELM CR	25	13.597	3.047	32.11	55	13.937	3.073	32.54
Elman	5	13.014	2.927	28.23	10	15.990	3.292	34.51
Jordan	5	11.482	2.698	26.52	15	11.810	2.742	28.00
ESN Ozturk	20	12.940	2.975	32.34	30	13.138	2.968	31.41
ESN Jaeger	35	12.863	2.960	31.92	65	12.515	2.891	29.87
ESN Ozturk CR	15	14.105	3.113	34.05	15	14.767	3.169	34.36
ESN Jaeger CR	35	13.370	3.016	31.96	85	13.973	3.014	31.28
GLM	–	15.331	3.219	29.16	–	15.654	3.227	27.26
GLM/PSO	–	18.858	3.567	41.43	–	19.590	3.646	42.26
Ensemble 1	–	12.052	2.872	30.99	–	11.882	2.855	30.66
Ensemble 2	–	11.720	2.825	30.08	–	11.530	2.805	29.70
Ensemble 3	**60**	**11.004**	**2.615**	**26.30**	5	11.267	2.678	25.73
Ensemble 4	15	11.127	2.640	27.80	**130**	**10.435**	**2.560**	**25.42**

The amount of results and scenarios addressed, allows to analyse many aspects regarding the errors in Table 2.

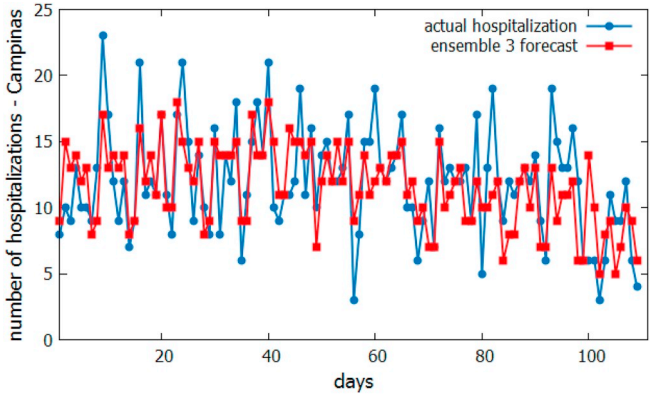

Fig. 9. Number of hospital admissions versus estimates by ensemble 3 for lag 3 - Campinas.

data were normalized in the same range. Then, we achieved the output responses, the data were renormalized to allow an analysis of the error in the original domain (Haykin et al., 2009).

The error metrics used in this study were the Mean Square Error (MSE), the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE), which formulations are given by 10, 11, 12:

$$MSE = \frac{1}{N_S}\sum_{t=1}^{N_S}(d_t - y_t)^2 \tag{10}$$

$$MAE = \frac{1}{N_S}\sum_{t=1}^{N_S}|d_t - y_t| \tag{11}$$

$$MAPE = \frac{1}{N_S}\sum_{t=1}^{N_S}\frac{d_t - y_t}{d_t} \tag{12}$$

being d_t the observed number of hospitalizations and y_t the output of the networks.

In cases in which the metrics indicate distinct models as the best, that with the lowest MSE was assumed as the best one (Siqueira et al., 2014).

All the results estimate the number of hospital admissions for up to seven days after exposure to PM_{10}, as suggests the literature (Polezer et al., 2018) (Tadano et al., 2016) (Li et al., 2015), because air pollution may affect population's health some days after the exposure. In this case, we applied the models to eight situations (Lag 0 to Lag 7), meaning:

- Lag 0 relates the concentration of PM_{10} from today with the health outcomes of the same day;
- Lag 1 relates the concentration of PM_{10} from today with the health outcomes of tomorrow;
- Lag 2 relates the concentration of PM_{10} from today with the health outcomes two days ahead, and son on.

The general performances of Campinas and São Paulo are presented in Tables 3 and 3, respectively. To neural models, the number of neurons in their hidden layer *NN* is also presented. The acronym RC represents regularization coefficient, and the ensembles are nominated as follows: Ensemble 1 - Average, Ensemble 2 - Median, Ensemble 3 - MLP, and Ensemble 4 - RBF. To all neural models and the GLM/PSO, 30 simulations were performed. Tables 2 and 3 present the best of these 30 cases, regarding the MSE (Tzanis et al., 2019) (Alimissis et al., 2018) (Ding et al., 2016).

Finally, the Friedman test was applied to verify if the results were significantly different. In all cases, the achieved p-values were very close to zero, indicating that a change in the model leads to distinct results.

To Campinas, among the results obtained to the estimation three days after exposure (lag 3), ensemble 3 with 175 neurons was the method with the lowest overall errors, considering the 3 aforementioned metrics. On the other hand, the estimate with worse result was achieved by the traditional GLM.

Comparing traditional GLM with GLM/PSO, the results showed PSO improved the performance, due to the better capability to perform global and local search simultaneously, in the sene of minimizing the output error.

Fig. 9 shows the comparison between actual number of hospital admissions (blue circles) and the results of the test set with the best predictor, the ensemble 3 method with 175 neurons, to lag 3 (red square) in Campinas. We need to highlight that countless factors may affect population's health, besides air quality, such as people's lifestyle, genetics, age and so on. With that in mind, and the poorness of data quality in developing countries, like Brazil, we may assume that the models gave a good estimate of air quality impact on population health.

Following, we presented Table 3 containing the errors to São Paulo estimates.

The best performance found was again favored to the ensemble using the MLP with 40 neurons, to all lags. Among the obtained results, the estimate performed to lag 7 achieved the lowest errors. Again, the superiority of the neural-based ensembles in comparison to the linear cases was highlighted.

Fig. 10 shows the comparison between the number of actual hospitalizations (blue circle) and the estimate performed to lag 7 by the ensemble based on the MLP (red square).

As a final analysis, it is clear that the comparison among all models was largely favorable to the ensemble combined by MLP, regarding the MSE. The reasons that can lead to this are diverse. We can firstly point the insertion of a large variety of input methods, such as ANN proposals and GLM. The diversity is important to form accurate outputs, as pointed by Matos Neto et al. (de Mattos Neto et al., 2014). This diversity is a preponderant factor to a good performance of ensembles.

However, it is important to discuss the use of the metrics addressed to analyse the comparative performance. Willmott and Matsuura (2005) advocate that the most relevant in assessing average model performances is the MAE, instead of the MSE. Considering this approach, in the case of São Paulo, the best result achieved is related to the ELM(CR) and still lag 7, being the Ensemble 3 the second best. To Campinas, the best results for all metrics were achieved by Ensemble 3.

It is evident that the use of ANN is more adequate to the studied cases in comparison to the linear models. This is important because the specialized literature of air pollution epidemiology is still very biased towards the use of classical GLM, especially the implementation in the R® software. In this sense, it is clear that there is still room for improvement of these methods.

Comparing each approach to create the ANN, the RBF achieved the best overall performances to the feedforward models. To the recurrent ones, the Jordan networks were the best to Campinas and the ESN from Ozturk et al. to São Paulo. This difference between scenarios is a strong indicative that the problem is complex to be solved and the more models available, the better may be the analysis.

Through the comparative analysis considering the ensembles, it is noticeable that the trainable methods, which use ANN as combiners, achieved the best performances in comparison to the non-trainable approaches (average and median). The main reason is the high capability of the neural models for mapping the data in a nonlinear manner and generalizing the results. As shown in Tables 2 and 3, the non-trainable ensembles can be even worse than the single models in some scenarios. Therefore, it seems to be clear that the ANN are more suitable to be used in this case.

The GLM/PSO proposal introduced in this study still needs to be evaluated in other scenarios, as well as more optimization algorithms should be tested. In addition, the spline methodology presents a degree of complexity often not understood by the researchers. Therefore, replace the spline by a simple deseasonalization process is a factor that favors the understanding of the method and can also lead to better

Table 3
Results for São Paulo.

Lag	0			1				
Model	NN	MSE	MAE	MAPE	NN	MSE	MAE	MAPE
MLP	5	936.366	25.313	26.51	5	958.059	25.389	26.82
RBF	25	917.457	25.334	26.44	15	958.924	25.533	27.59
ELM	15	949.859	25.784	27.20	70	984.680	25.583	25.81
ELM CR	55	911.078	24.800	24.96	40	937.496	24.793	25.27
Elman	5	2793.758	44.022	50.90	5	1254.940	29.325	32.51
Jordan	5	968.469	25.615	26.80	5	1005.852	26.227	27.78
ESN Ozturk	50	934.164	25.203	26.12	25	963.843	25.586	26.68
ESN Jaeger	20	937.153	25.248	26.60	40	1071.072	26.764	27.18
ESN Ozturk CR	25	915.959	24.854	25.39	65	913.339	24.163	24.44
ESN Jaeger CR	35	916.445	24.784	25.36	30	931.899	25.294	26.38
GLM	–	1668.927	34.316	35.83	–	1711.703	34.833	35.85
GLM/PSO	–	1429.927	30.993	35.48	–	1430.529	31.163	35.29
Ensemble 1	–	961.364	25.955	28.284	–	973.253	25.840	27.73
Ensemble 2	–	954.122	25.770	27.756	–	983.246	25.864	27.39
Ensemble 3	**30**	**870.607**	**24.254**	**24.92**	**25**	**904.646**	**24.580**	**25.62**
Ensemble 4	5	998.282	25.730	27.04	5	1114.375	26.896	27.57
Lag	2			3				
MLP	5	964.663	25.823	27.22	5	979.703	25.659	25.98
RBF	60	960.612	24.963	25.10	115	976.647	25.110	25.48
ELM	30	968.737	25.965	26.49	30	983.791	25.666	26.15
ELM CR	20	961.716	25.640	25.92	30	960.857	25.128	25.71
Elman	5	1318.999	30.294	33.48	5	1219.160	28.807	31.76
Jordan	5	994.172	26.177	27.85	5	1043.520	26.667	28.58
ESN Ozturk	20	946.243	25.349	26.49	40	1032.186	26.324	27.59
ESN Jaeger	25	980.076	25.751	26.41	20	982.565	25.817	26.53
ESN Ozturk CR	25	955.644	25.428	26.12	60	939.485	24.415	24.64
ESN Jaeger CR	20	952.176	25.448	26.40	35	955.736	25.624	25.90
GLM	–	1606.712	33.619	35.52	–	1697.937	34.720	35.71
GLM/PSO	–	1449.639	31.003	35.33	–	1416.267	30.552	35.07
Ensemble 1	–	991.241	25.974	27.89	–	975.812	25.450	27.26
Ensemble 2	–	1000.730	26.008	27.58	–	988.877	25.484	26.91
Ensemble 3	**25**	**904.298**	**24.572**	**25.03**	**50**	**899.943**	**24.752**	**25.83**
Ensemble 4	5	1192.876	27.815	28.54	5	1157.295	27.719	27.786
Lag	4			5				
MLP	5	968.378	26.142	27.17	5	1012.306	26.468	27.89
RBF	55	965.352	25.171	25.61	145	931.952	24.979	25.04
ELM	25	983.716	25.890	26.58	20	974.059	25.585	27.35
ELM CR	35	968.608	25.269	25.435	30	967.783	25.423	26.28
Elman	5	2568.163	42.021	48.80	5	1301.298	29.881	32.99
Jordan	5	1025.605	26.631	27.98	5	1057.722	26.876	27.64
ESN Ozturk	45	999.682	25.951	26.89	20	987.100	26.079	27.74
ESN Jaeger	40	1000.702	25.979	26.19	30	973.066	25.322	26.32
ESN Ozturk CR	50	947.300	24.561	24.94	30	940.991	25.243	26.31
ESN Jaeger CR	25	964.764	25.636	26.50	35	944.669	25.185	26.33
GLM	–	1650.754	34.117	35.54	–	1591.655	33.319	35.22
GLM/PSO	–	1435.435	31.104	35.52	–	1421.780	31.173	35.36
Ensemble 1	–	1002.979	26.210	28.37	–	968.109	25.619	27.71
Ensemble 2	–	1003.977	26.165	27.95	–	974.294	25.599	27.33
Ensemble 3	**25**	**908.431**	**24.394**	**24.62**	**40**	**900.408**	**24.714**	**25.91**
Ensemble 4	5	1019.277	26.146	27.22	5	1131.680	27.205	28.09
Lag	6			7				
MLP	5	979.655	25.790	27.16	5	934.197	25.587	27.35
RBF	25	986.427	26.095	28.28	25	952.152	25.634	27.70
ELM	25	954.977	25.688	27.32	25	897.338	24.801	26.21
ELM CR	30	877.917	24.306	25.00	55	875.590	24.023	24.58
Elman	5	2555.438	41.611	48.87	5	1186.482	28.610	31.48
Jordan	5	1021.505	26.558	28.62	5	983.723	25.992	27.60
ESN Ozturk	25	939.547	25.130	26.80	65	920.228	24.891	25.47
ESN Jaeger	40	945.313	25.370	26.41	25	984.248	25.599	26.14
ESN Ozturk CR	40	906.183	24.439	25.32	50	866.800	23.769	34.65
ESN Jaeger CR	50	927.146	24.851	25.52	55	881.039	24.191	24.97
GLM	–	1524.647	32.324	35.02	–	1649.185	34.166	35.27
GLM/PSO	–	1480.968	31.562	36.46	–	1434.060	31.104	35.69
Ensemble 1	–	983.774	26.193	28.94	–	918.631	25.258	27.29
Ensemble 2	–	975.961	26.039	28.43	–	919.822	25.223	26.89
Ensemble 3	**40**	**890.179**	**24.424**	**25.21**	**40**	**855.119**	**24.037**	**25.79**
Ensemble 4	5	1139.652	27.011	28.25	5	1070.765	26.845	27.68

performance. The general analysis shows that the GLM/PSO is a competitive candidate in this area.

As for the number of lags ideal to each case, it is difficult to assert an exact value if all models are considered, since for each one there was a discrepancy in the best results. This is not a surprising factor and it shows how complex the problem is. The same observation applies to the number of neurons in the middle layer of neural networks.

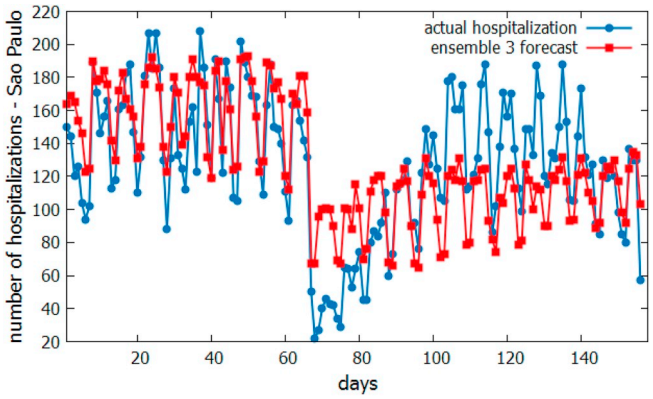

Fig. 10. Number of hospitalizations for respiratory diseases versus prediction ensemble 3 with neurons 175 for lag 7 - São Paulo.

6. Conclusion

The general analysis revealed some important points. First, the GLM/PSO won the classical proposal to São Paulo. However, it is worth noting that the general results strongly indicate that linear models can not achieve lower errors than nonlinear performances. Second, in this same scenario, the ESN from Ozturk et al. was able to overcome all the classic recurring methodologies. And, the most important result is that, even with a high dispersion, MLP-based neural ensemble was the best in both cases.

The new approach of GLM and the use of ANN and ensembles showed significant improvements to classical GLM. This research opens the horizon to the methodologies that may be applied to assess health adverse effects of air pollution, as it is a complex problem to be solved. Expanding for example, possibilities of researches at areas with small amount of data and improving predictions. We need to highlight that the present research showed the proposed methodologies has succeeded on estimating air quality impact on population health, even with the assumption that measured air quality is representative of the whole city and the citizens exposure. Then, we recommend that the models should be tested to more representative data to assure their performance.

The main difficulty was to capture peaks of hospitalizations, that must be considered with more caution in future researches, however, the results of ANN and ensembles were better than those of GLM. There are an important difference between scenarios, which is a strong indicative that the problem is complex to be solved. In this sense, the more models available, the better may be the analysis.

We understand that the best action to improve air quality is reducing emissions, however it is not possible in a short term. Then, due to the good performance of ANN and ensembles, they could be used to estimate hospital admissions and alert government and hospitals for possible increases of hospitalizations due to high air pollution episodes or adverse meteorological conditions.

Also, developing countries, like Brazil, has a lack of information about air quality, due to financial resources, making it difficult sometimes to estimate air quality impacts on human health using statistical regressions. Then, universal approximators like ANN may allow studies of air quality health impacts to be done, as it can find relationship between variables even with small databases.

As a future perspectives to improve and spread the available methodologies used to assess health air pollution impacts, it is proposed to use other optimization algorithms for GLM, such as genetic, differential or immuno-inspired evolution. In the case of neural networks, it is worth discussing the use of other ensemble combiners, such as ELM itself. Using other weather variables besides temperature and relative humidity, other types of pollutants other than PM_{10}, may also be alternatives. Deep learning techniques depending on the number of data available, can also be applied to the problem.

Acknowledgment

The authors would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico for the financial support, under process number 405580/2018-5.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2019.104567.

References

Alimissis, A., Philippopoulos, K., Tzanis, C., Deligiorgi, D., 2018. Spatial estimation of urban air pollution with the use of artificial neural network models. Atmos. Environ. 191, 205–213. https://doi.org/10.1016/j.atmosenv.2018.07.058.

Boccato, L., Lopes, A., Attux, R., Von Zuben, F.J., 2011. An echo state network architecture based on volterra filtering and pca with application to the channel equalization problem. In: Neural Networks (IJCNN), the 2011 International Joint Conference on. IEEE, pp. 580–587. https://doi.org/10.1109/IJCNN.2011.6033273.

Cabaneros, S.M.S., Calautit, J.K., Hughes, B.R., 2019. A review of artificial neural network models for ambient air pollution prediction. Environ. Model. Softw 119, 285–304. https://doi.org/10.1016/j.envsoft.2019.06.014.

CETESB, 2010. Companhia de Tecnologia de Saneamento Ambiental, Air quality of Sao Paulo state 2007, 2008 and 2009 (in Portuguese: Qualidade do ar no estado de São Paulo 2007, 2008 e 2009). https://cetesb.sp.gov.br/ar/publicacoes-relatorios/. (Accessed 27 January 2019).

CETESB, 2018. Companhia de Tecnologia de Saneamento Ambiental, Air quality of Sao Paulo state 2018 (in Portuguese: Qualidade do ar no estado de São Paulo 2018). https://cetesb.sp.gov.br/ar/publicacoes-relatorios/. (Accessed 27 August 2019).

Chelani, A.B., Rao, C.C., Phadke, K., Hasan, M., 2002. Prediction of sulphur dioxide concentration using artificial neural networks. Environ. Model. Softw 17 (2), 159–166. https://doi.org/10.1016/S1364-8152(01)00061-5.

Coman, A., Ionescu, A., Candau, Y., 2008. Hourly ozone prediction for a 24-h horizon using neural networks. Environ. Model. Softw 23 (12), 1407–1421. https://doi.org/10.1016/j.envsoft.2008.04.004.

DATASUS, 2010. Departamento de Informática do Sistema Único de Saúde, Health informations (in Portuguese: Informações de saúde). https://www2.datasus.gov.br/DATASUS/index.php?/area=0701&item=201&acao=11. (Accessed 27 January 2019).

de Campinas, Prefeitura, 2014. The City - Geographic Data (In Portuguese: A Cidade - Dados Geográficos). https://www.campinas.sp.gov.br/governo/seplama//dados-do-municipio/cidade/. (Accessed 2 November 2018).

de Mattos Neto, P.S., Madeiro, F., Ferreira, T.A., Cavalcanti, G.D., 2014. Hybrid intelligent system for air quality forecasting using phase adjustment. Eng. Appl. Artif. Intell. 32 (0), 185–191. https://doi.org/10.1016/j.engappai.2014.03.010.

Ding, W., Zhang, J., Leung, Y., 2016. Prediction of air pollutant concentration based on sparse response back-propagation training feedforward neural networks. Environ. Sci. Pollut. Control Ser. 23 (19), 19481–19494. https://doi.org/10.1007/s11356-016-7149-4.

Dobson, A.J., Barnett, A.G., 2008. An Introduction to Generalized Linear Models. Chapman and Hall/CRC.

Dobson, A., Barnett, A., 2008. An Introduction to Generalized Linear Models. Chapman and Hall/CRC.

Du, K.-L., Swamy, M., 2014. Radial basis function networks. In: Neural Networks and Statistical Learning. Springer, pp. 299–335.

Eberhart, R., Kennedy, J., 1995. A new optimizer using particle swarm theory. In: Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on. IEEE, pp. 39–43. https://doi.org/10.1109/MHS.1995.494215.

Elman, J.L., 1990. Finding structure in time. Cogn. Sci. 14 (2), 179–211.

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

P. R. A. Firmino, P. S. de Mattos Neto, T. A. Ferreira, Correcting and combining time series forecasters, Neural Netw. 50 (0) 1 – 11. doi:https://doi.org/10.1016/j.neunet.2013.10.008.

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

Haykin, S.S., Haykin, S.S., Haykin, S.S., Haykin, S.S., 2009. Neural Networks and Learning Machines, vol. 3. Pearson Upper, Saddle River, NJ, USA.

Huang, G.-B., Chen, L., Siew, C.K., et al., 2006. Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Trans. Neural Netw. 17 (4), 879–892. https://doi.org/10.1109/TNN.2006.875977.

IBGE, 2019. Instituto Brasileiro de Geografia e Estatística, Estimates of living population at cities and federation units (in Portuguese: Estimativas da população residente para os municípios e para as unidades da federação). https://www.ibge.gov.br/estatisticas-novoportal/sociais/populacao/9103-estimativas-de-populacao.html?=&t=resultados. (Accessed 27 January 2019).

Jaeger, H., 2001. The "Echo State" Approach to Analysing and Training Recurrent Neural Networks-With an Erratum Note, vol. 148. German National Research Center for Information Technology GMD Technical Report, Bonn, Germany, p. 13, 34.

Kassomenos, P., Petrakis, M., Sarigiannis, D., Gotti, A., Karakitsios, S., 2011. Identifying the contribution of physical and chemical stressors to the daily number of hospital admissions implementing an artificial neural network model. Air Qual. Atmos. Health 4 (3–4), 263–272. https://doi.org/10.1007/s11869-011-0139-2.

Kim, J., Lee, J.Y., 2019. Synoptic approach to evaluate the effect of temperature on pediatric respiratory disease-related hospitalization in Seoul, Korea. Environ. Res. 178, 108650. https://doi.org/10.1016/j.envres.2019.108650.

Kuan, C., 1989. Estimation of Neural Network Models. University Microfilms Int./UMI. https://books.google.com.br/books?id=SOQUnQEACAAJ.

Lauret, P., Heymes, F., Aprin, L., Johannet, A., 2016. Atmospheric dispersion modeling using artificial neural network based cellular automata. Environ. Model. Softw 85, 56–69. https://doi.org/10.1016/j.envsoft.2016.08.001.

Li, Y., Ma, Z., Zheng, C., Shang, Y., 2015. Ambient temperature enhanced acute cardiovascular-respiratory mortality effects of PM2.5 in beijing, China. Int. J. Biometeorol. 59 (12), 1761–1770. https://doi.org/10.1007/s00484-015-0984-z.

Markatos, N., 1985. Numerical Methods for Engineers with Personal Computer Applications: Steven C. Chapra and Raymond P. Canale. mcgraw-hill, new york xiv+ 570 pp (1987).

Matsuzaki, T., 2017. Understanding the Basis of Glm Regression. https://tsmatz.wordpress.com/2017/08//30/glm-Ωregression-logistic-poisson-gaussian-gamma-tutorial-with-r.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. Chapman and Hall/CRC, Washington, D.C.

Mishra, D., Goyal, P., Upadhyay, A., 2015. Artificial intelligence based approach to forecast PM2.5 during haze episodes: a case study of Delhi, India. Atmos. Environ. 102, 239–248. https://doi.org/10.1016/j.atmosenv.2014.11.050.

Neuhaus, J., McCulloch, C., 2011. Generalized linear models. Wiley Interdiscipl. Rev.: Comput. Stat. 3 (5), 407–413. https://doi.org/10.1002/wics.175.

Ozturk, M.C., Xu, D., Príncipe, J.C., 2007. Analysis and design of echo state networks. Neural Comput. 19 (1), 111–138. https://doi.org/10.1162/neco.2007.19.1.111.

Pan, R., Gao, J., Wang, X., Bai, L., Wei, Q., Yi, W., Xu, Z., Duan, J., Cheng, Q., Zhang, Y., et al., 2019. Impacts of exposure to humidex on the risk of childhood asthma hospitalizations in Hefei, China: effect modification by gender and age. Sci. Total Environ. 691, 296–305. https://doi.org/10.1016/j.scitotenv.2019.07.026.

Paula, G.A., 2004. Modelos de regressão: com apoio computacional. IME-USP São Paulo. https://www.ime.usp.br/~giapaula/texto_2013.pdf.

Perrone, M.P., Cooper, L.N., 1992. When Networks Disagree: Ensemble Methods for Hybrid Neural Networks, Tech. Rep. Brown Univ Providence Ri Inst fro Brain and Neural Systems.

Polezer, G., Tadano, Y.S., Siqueira, H.V., Godoi, A.F., Yamamoto, C.I., de André, P.A., Pauliquevis, T., de Fatima Andrade, M., Oliveira, A., Saldiva, P.H., et al., 2018. Assessing the impact of PM2.5 on respiratory disease using artificial neural networks. Environ. Pollut. 235, 394–403. https://doi.org/10.1016/j.envpol.2017.12.111.

Powell, M.J.D., 1987. Algorithms for Approximation. Clarendon Press, New York, NY, USA, pp. 143–167. Ch. Radial Basis Functions for Multivariable Interpolation: A Review. http://dl.acm.org/citation.cfm?id=48424.48433.

Ren, G., Cao, Y., Wen, S., Huang, T., Zeng, Z., 2018. A modified elman neural network with a new learning rate scheme. Neurocomputing 286, 11–18. https://doi.org/10.1016/j.neucom.2018.01.046.

Siqueira, H., Luna, I., 2019. Performance comparison of feedforward neural networks applied to stream flow series forecasting. J. MESA 10 (1), 41–53.

Siqueira, H., Boccato, L., Attux, R., Lyra, C., 2014. Unorganized machines for seasonal streamflow series forecasting. Int. J. Neural Syst. 24 (03), 1430009. https://doi.org/10.1142/S0129065714300095.

Siqueira, H., Boccato, L., Luna, I., Attux, R., Lyra, C., 2018. Performance analysis of unorganized machines in streamflow forecasting of brazilian plants. Appl. Soft Comput. 68, 494–506. https://doi.org/10.1016/j.asoc.2018.04.007.

Spark, Weather, 2018. Mean meteorological conditions of Campinas and region (in Portuguese: Condições meteorológicas médias de Campinas e região). https://pt.weatherspark.com/y/30197/Clima-caracter%C3%ADstico-em-Campinas-e-Regi/~ao-/Brasil-durante-o-ano. (Accessed 3 November 2018).

Sundaram, N.M., Sivanandam, S., Subha, R., 2016. Elman neural network mortality predictor for prediction of mortality due to pollution. Int. J. Appl. Eng. Res. 11 (3), 1835–1840.

A. M. Turing, Intelligent Machinery, a Heretical Theory, the Turing Test: Verbal Behavior as the Hallmark of Intelligence 105.

Tadano, Y.S., Siqueira, H.V., Alves, T.A., 2016. Unorganized machines to predict hospital admissions for respiratory diseases. In: Computational Intelligence (LA-CCI), 2016. IEEE Latin American Conference on, pp. 1–6. https://doi.org/10.1109/LA-CCI.2016.7885699.

Tadano, Y.S., Ugaya, C.M.L., Franco, A.T., 2012. Methodology to assess air pollution impact on human health using the generalized linear model with Poisson regression. In: Khare, M. (Ed.), Air Pollution-Monitoring, Modelling and Health. InTech, pp. 281–304. https://doi.org/10.5772/33385.

Tzanis, C.G., Alimissis, A., Philippopoulos, K., Deligiorgi, D., 2019. Applying linear and nonlinear models for the estimation of particulate matter variability. Environ. Pollut. 246, 89–98. https://doi.org/10.1016/j.envpol.2018.11.080.

Vanos, J.K., Hebbern, C., Cakmak, S., 2014. Risk assessment for cardiovascular and respiratory mortality due to air pollution and synoptic meteorology in 10 Canadian cities. Environ. Pollut. 185, 322–332. https://doi.org/10.1016/j.envpol.2013.11.007.

Wang, Q., Liu, Y., Pan, X., 2008. Atmosphere pollutants and mortality rate of respiratory diseases in Beijing. Sci. Total Environ. 391 (1), 143–148. https://doi.org/10.1016/j.scitotenv.2007.10.058.

WHO, 2017. Evolution of Who Air Quality Guidelines: Past, Present and Future. Who Regional Office for Europe, Copenhagen, 2017. https://www.euro.who.int/_data/assets/pdf_file/0019/331660/Evolution-air-quality.pdf/?ua=1.

Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. Clim. Res. 30 (1), 79–82. https://doi.org/10.3354/cr030079.

Yildiz, I.B., Jaeger, H., Kiebel, S.J., 2012. Re-visiting the echo state property. Neural Netw. 35 (Suppl. C), 1–9. https://doi.org/10.1016/j.neunet.2012.07.005.