

Article

Unorganized Machines to Estimate the Number of Hospital Admissions Due to Respiratory Diseases Caused by PM₁₀ Concentration

Yara de Souza Tadano ^{1,*}, Eduardo Tadeu Bacalhau ², Luciana Casacio ², Erickson Puchta ³, Thomas Siqueira Pereira ⁴, Thiago Antonini Alves ⁴, Cássia Maria Lie Ugaya ⁵ and Hugo Valadares Siqueira ³

¹ Department of Mathematics, Federal University of Technology, 330 Doutor Washington Subtil Chueire Street, Ponta Grossa 84017-220, PR, Brazil

² Center for Marine Studies, Pontal do Paraná Campus, Federal University of Paraná, Beira-mar Avenue, P.O. Box 61, Pontal do Paraná 83255-976, PR, Brazil; bacalhau@ufpr.br (E.T.B.); lucianacasacio@ufpr.br (L.C.)

³ Department of Electric Engineering, Federal University of Technology, 330 Doutor Washington Subtil Chueire Street, Ponta Grossa 84017-220, PR, Brazil; ericksonpuchta@gmail.com (E.P.); hugosiqueira@utfpr.edu.br (H.V.S.)

⁴ Department of Mechanical Engineering, Federal University of Technology, 330 Doutor Washington Subtil Chueire Street, Ponta Grossa 84017-220, PR, Brazil; thomassiqueira427@gmail.com (T.S.P.); antonini@utfpr.edu.br (T.A.A.)

⁵ Department of Mechanical, Federal University of Technology, CNPq Fellow, 5000 Dep. Heitor Alencar Furtado Street, Curitiba 81280-340, PR, Brazil; cassiaugaya@utfpr.edu.br

* Correspondence: yarataadano@utfpr.edu.br



Citation: Tadano, Y.d.S.; Bacalhau, E.T.; Casacio, L.; Puchta, E.; Pereira, T.S.; Antonini Alves, T.; Ugaya, C.M.L.; Siqueira, H.V. Unorganized Machines to Estimate the Number of Hospital Admissions Due to Respiratory Diseases Caused by PM₁₀ Concentration. *Atmosphere* **2021**, *12*, 1345. <https://doi.org/10.3390/atmos12101345>

Academic Editors: Hsiao-Chi Chuang and Alina Barbulescu

Received: 17 July 2021

Accepted: 6 October 2021

Published: 14 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: The particulate matter PM₁₀ concentrations have been impacting hospital admissions due to respiratory diseases. The air pollution studies seek to understand how this pollutant affects the health system. Since prediction involves several variables, any disparity causes a disturbance in the overall system, increasing the difficulty of the models' development. Due to the complex nonlinear behavior of the problem and their influencing factors, Artificial Neural Networks are attractive approaches for solving estimations problems. This paper explores two neural network architectures denoted unorganized machines: the echo state networks and the extreme learning machines. Beyond the standard forms, models variations are also proposed: the regularization parameter (RP) to increase the generalization capability, and the Volterra filter to explore nonlinear patterns of the hidden layers. To evaluate the proposed models' performance for the hospital admissions estimation by respiratory diseases, three cities of São Paulo state, Brazil: Cubatão, Campinas and São Paulo, are investigated. Numerical results show the standard models' superior performance for most scenarios. Nevertheless, considering divergent intensity in hospital admissions, the RP models present the best results in terms of data dispersion. Finally, an overall analysis highlights the models' efficiency to assist the hospital admissions management during high air pollution episodes.

Keywords: PM₁₀; health risks; extreme learning machine; echo state network; neural networks



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

World Health Organization (WHO) estimates that 91% of the world's population lives in places where air pollution levels exceed the advised limits. This exposure has as a consequence 4.2 million deaths per year due to stroke, heart disease, lung cancer and chronic respiratory illness [1].

In the last decades, the air pollution consequences in the environment and health have been the subject of deep researches [2–4], including the relation between air pollution and human health [5–8] and, specifically, the study of particulate matter (PM) impacts on the respiratory diseases [9–11]. The public health system is currently the main concern for the global governance majority, receiving huge money investments and boosting researches in

operational areas. Therefore, several works have been applied to develop mathematical models to improve predicting the diseases caused by PM air concentration.

Generalized Linear Models (GLM) [10–14] and Generalized Additive Models (GAM) [15,16] are statistical regression models usually used to assess air pollution consequences on human health. However, a minimum of data is required to assure that regression models will be able to capture the relationship between the inputs (predictors) and the output (response variable) [17]. For developing countries, as lack of data is a reality, solving the problem using regression models is challenging [18]. For this reason, other models and methods have been applied; since the problem can be seen as a nonlinear mapping task, the Artificial Neural Networks (ANN) approach is the most attractive approach for solving estimation problems. The ANN have been used to solve air pollution mapping tasks [19–22], and they have become increasingly popular over the past decade for predicting the air pollutant's impact on human health [10,17,18,23–25]. Araujo et al. [17] and Kassomenos et al. [24] have shown that the ANN had better performance than linear approaches like the GLM when dealing with nonlinear mapping problems. In this context, Tadano et al. [26] proposed to use two models, known as Unorganized Machines (UM): the echo state networks (ESN) and the extreme learning machines (ELM), to predict hospital admissions. Based on this work, this paper presents a full extension of these models, adding several neural networks variations applied to an enlarged and updated set of instances.

ELM and ESN are ANN architectures used to deal with static nonlinear mapping problems, and are reliable when applied to multiclass classification and, mainly, time series forecasting [27–31]. Thus, the main contribution of this research is an epistemological study that predicts the impact of PM₁₀ (particulate matter with an aerodynamic diameter less than 10 µm) daily mean concentrations on hospital admissions due to respiratory diseases using versions of the UM: the addition of regularization parameter applied to increase the generalization capability of the models [32] and the use of the Volterra filter to capture nonlinear patterns of the neural information [33]. To evaluate the performance of the proposed methods, three cities from São Paulo State, Brazil (Campinas, Cubatão and São Paulo city) were considered.

Based on the overall analysis produced, we expect to understand how air pollution affects the health system, especially during global sanitary crises scenarios, avoiding hospital collapse.

This work is organized as follows: Section 2 presents the ELM and ESN standard models, the regularization parameter and nonlinear output layer strategies; Section 3 describes the addressed databases; Section 4 shows the computational results and critical analysis regarding the models' performances; Section 5 presents the main conclusions and future works.

2. Unorganized Machines

Unorganized machines are a designation used as a general term to classify the modern neural network paradigms that unify two kinds of ANN: the echo state networks (ESNs) and the extreme learning machines (ELMs) [27].

In this work, these two architectures are employed to predict the hospital admissions due to respiratory diseases caused by air pollution. Moreover, other models based on the variations and extensions of these models are used [33,34].

2.1. Extreme Learning Machines

The extreme learning machine (ELM) is a feedforward neural network composed of a single hidden layer, similar to the structure of multilayer perceptron (MLP) [28]. Figure 1 illustrates the architecture.

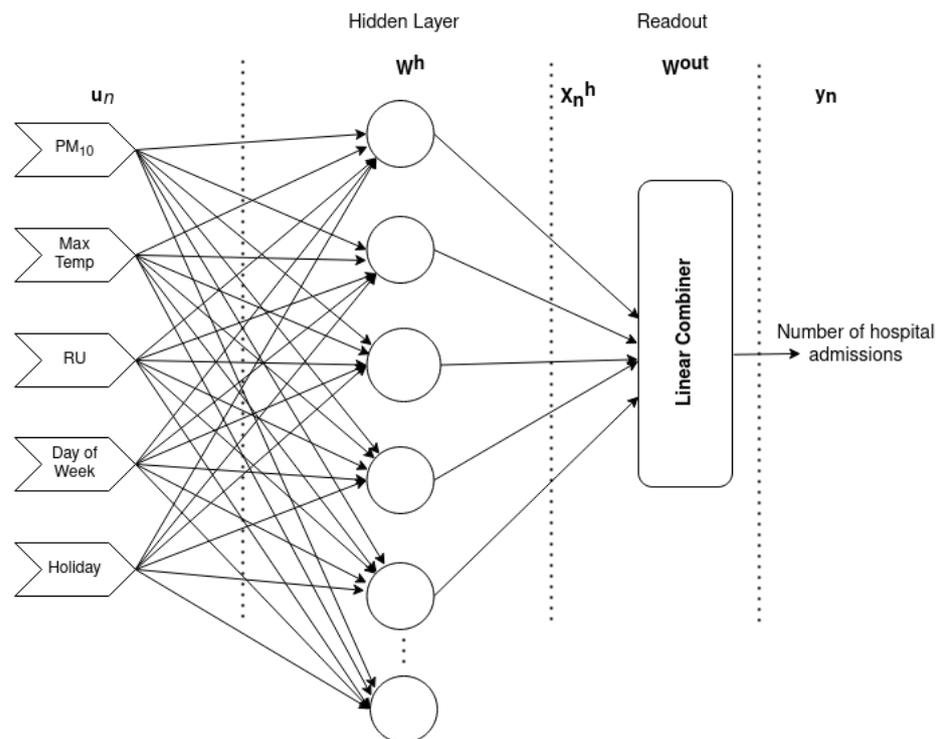


Figure 1. Extreme Learning Machine.

According to Figure 1, the vector \mathbf{u}_n represents all input information: PM₁₀ concentration; relative humidity; ambient temperature; the different weekdays; and holidays. This vector \mathbf{u}_n is associated with the matrix \mathbf{W}^h through weights of the hidden layer that can be randomly determined. The unique output layer (readout) \mathbf{W}^{out} is composed of parameters of a linear combiner that are calculated using the Moore-Penrose generalized inverse operator which shall be defined below. Finally, similar to a single-hidden layer multilayer perceptron (MLP), the ELM is also a single hidden layer feedforward neural network, being y_n the output information that indicates the number of hospital admissions.

The activation of the artificial neurons within the hidden layer are given by Equation (1):

$$\mathbf{x}_n^h = f_n^h(\mathbf{W}^h \mathbf{u}_n + \mathbf{b}), \tag{1}$$

being $\mathbf{u}_n = [u_n, u_{n-1}, \dots, u_{n-K-1}]^T$ the vector that contains the K input signals, $\mathbf{W}^h \in \mathbb{R}^{N \times K}$ the linear input coefficients, \mathbf{b} the vector that represents the biases of the hidden units and $f^h(\cdot) = (f_1^h(\cdot), f_2^h(\cdot), \dots, f_N^h(\cdot))$ the activation functions of the hidden neurons. Then, Equation (2) presents the network outputs calculation:

$$y_n = \mathbf{W}^{out} \mathbf{x}_n^h, \tag{2}$$

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

The output layer (readout) adjustment is the main advantage of ELM models. This strategy is applied only once, considering the error signal [35,36]. Moreover, in dissonance with the traditional feedforward neural networks, when the intermediate activation functions are continuously differentiable, these models can choose the weights of the hidden layer randomly [36–38]. Huang et al. demonstrate that ELMs are universal approximators [39].

These structures are composed of a simple training process, mainly requiring the calculation of the parameters of a linear combiner using the Moore-Penrose generalized inverse operator, as in Equation (3) [36,37,40,41]:

$$\mathbf{W}^{out} = (\mathbf{X}_h^T \mathbf{X}_h)^{-1} \mathbf{X}_h^T \mathbf{d}, \tag{3}$$

where $\mathbf{X}_h \in \mathbb{R}^{T_s \times N}$ is the matrix composed of the intermediate layer outputs and T_s is the training sample numbers, $(\mathbf{X}_h^T \mathbf{X}_h)^{-1} \mathbf{X}_h^T$ is the pseudoinverse of \mathbf{X}_h and $\mathbf{d} \in \mathbb{R}^{T_s \times 1}$ is the vector composed of desired outputs.

2.2. Echo State Networks

Echo state networks (ESN) are recurrent neural models known by an effortless training process: the dynamical reservoir (intermediate layer) is fixed, i.e., there is no iterative adjustment. In this sense, the synaptic weights of the reservoir do not use the error function derivatives. Thus, only the output layer is effectively adapted [42]. The adaptation process applies a linear regression scheme similar to the ELM training process, considering that a linear combiner is often applied to the output layer. The neural network structure of ESN can be seen as a general case of ELM because the reservoir presents recurrent loops. Figure 2 illustrates the structure.

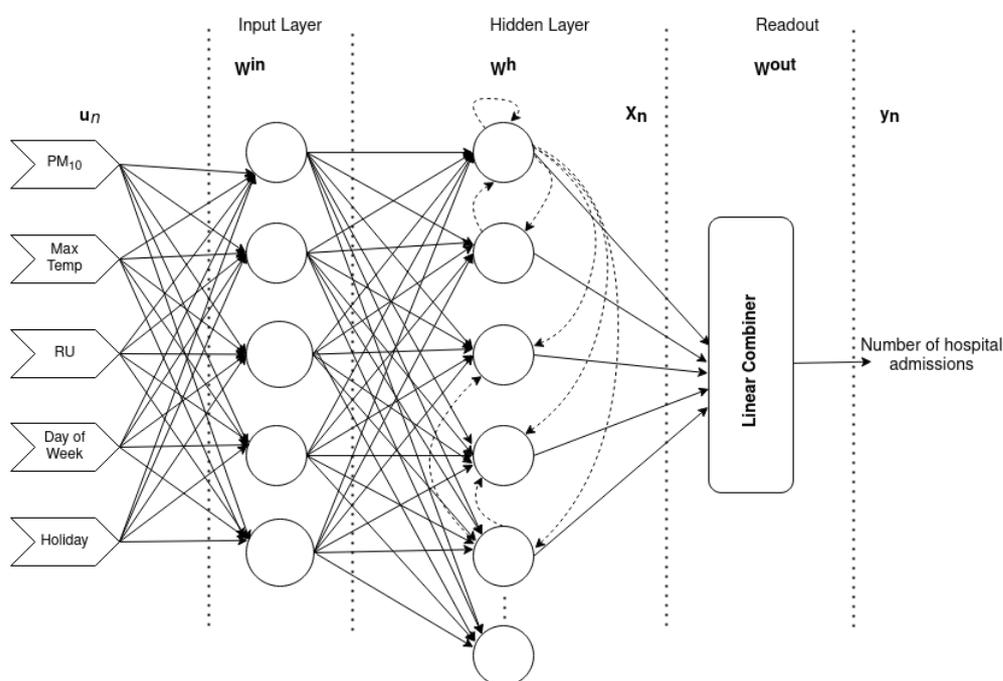


Figure 2. Echo state networks.

Figure 2 shows that the network structure is slightly similar to the ELM model presented in Figure 1, except by the additional input layer (\mathbf{W}^{in}), defined as a linear matrix, and feedback loops in the intermediate layer (hidden layer).

Equation (4) expresses the activation of the internal neurons. This activation represents the network states which are influenced by the previous state and the present input:

$$\mathbf{x}_{n+1} = f(\mathbf{W}^{in} \mathbf{u}_{n+1} + \mathbf{W} \mathbf{x}_n), \tag{4}$$

where $f(\cdot) = (f_1(\cdot), f_2(\cdot), \dots, f_N(\cdot))$ gives the activation functions of all neurons within the reservoir, $\mathbf{W}^{in} \in \mathbb{R}^{N \times K}$ is the input weight matrix and $\mathbf{W} \in \mathbb{R}^{N \times N}$ is the recurrent weight matrix.

The linear combinations of the reservoir signals produce the ESN outputs by (5):

$$\mathbf{y}_{n+1} = \mathbf{W}^{out} \mathbf{x}_{n+1}, \tag{5}$$

where $\mathbf{W}^{out} \in \mathbb{R}^{O \times N}$ is the output weight matrix, and O the number of outputs. The parameters of the \mathbf{W}^{out} are determined by Moore-Penrose generalized inverse described in Section 2.1.

Fundamentally, the network model, besides a stable behavior, should present an internal memory that preserves the input signals history formed in the dynamical reservoir [29,35,43]. Both features are contemplated by echo state property (ESP) [29,35,43].

Jaeger et al. suggest in [29] to simplify the weight matrix W , denoting w_{ij} as 0, 0.4 and -0.4 values with probabilities 0.95, 0.025 and 0.025, respectively. On the other hand, Ozturk et al. (2006) suggest a new design for the dynamical reservoir [44] that considers eigenvalues uniformly spreading in the weight matrix. Both approaches are applied in this work.

Having described the unorganized machines in the standard forms, the following subsections describe the variations and extensions which design structures of new models also applied to the proposed problem.

2.3. Regularization Parameter

Primarily proposed by Huang et al. (2011), the regularization strategy aims to improve the model's generalization capability, inducing the solutions obtained by a parameter applied to the Mean Square Error (MSE) cost function. The parameter C is chosen from a validation set of samples, assuming $C = 2^\lambda$, with λ discretized in the interval $[-25, 26]$ [32]. The strategy is performed during the interactive process, where all parameters are tested, and only one is selected according to the best MSE validation, via Expression (6):

$$\mathbf{W}^{out} = \left(\frac{\mathbf{I}}{C} + \mathbf{X}_h^T \mathbf{X}_h \right)^{-1} \mathbf{X}_h^T \mathbf{d}, \quad (6)$$

being C the regularization parameter and \mathbf{I} the identity matrix.

Trying to improve generalization capability given by the parameter C , Kulaif et al. (2013) developed a local search, denoted golden search, to determine better values for the parameter C . The strategy is grounded in two main concepts: significant modifications are obtained in the final solutions if any small parameter variations occur; the function given by each small interval associated with the parameter C and the validation error shall be supposedly quasi-convex [45]. This strategy is also applied in this work.

2.4. Nonlinear Output Layer

Boccatto et al. (2011) proposed a variation of nonlinear output layer in ESNs, the Volterra filtering structure [46]. The main concern is to prove the linear dependence between the dynamical echo states, preserving the training process simplicity for the networks. The output signals can be computed through linear combinations of polynomial terms, as in Equation (7) [27]:

$$y_{i,n} = h^0 + \sum_{p=1}^M h_p^1 x_{p,n} + \sum_{p=1}^M \sum_{q=1}^M h_{p,q}^2 x_{p,n} x_{q,n} + \sum_{p=1}^M \sum_{q=1}^M \sum_{r=1}^M h_{p,q,r}^3 x_{p,n} x_{q,n} x_{r,n} + \dots, \quad (7)$$

where $x_{i,n}$ is the output of the i -th neuron of the reservoir (or the i -th echo state) at n -th time instant, h^m the linear combiner coefficient with $m = 1, \dots, M$, and M the polynomial expansion order.

Similar to Equation (3), the training process simplicity is preserved due to the linear dependence of the outputs regarding the filter parameters. In terms of least squares, Equation (7) guarantee the closed-form solution, allowing the Moore-Penrose inverse operation [47].

However, according to Boccatto et al. (2011), the application of a Volterra filter might have as consequence the uncontrollable growth of free parameters and inputs numbers. To prevent these problems, a compression technique known as Principal Component Analysis (PCA) must be applied. Interestingly, the use of PCA is also suitable to avoid the redundancy between echo states [29,48]. In recent years, Chen et al. extended this idea to the ELMs, considering the same premises of the former work [48,49].

All parameters associated with the proposed models: the number of neurons, Volterra Filter orders, the weight values, and the number of simulations, shall be described in Section 4.

3. Case Studies

To evaluate the approach, three cities of São Paulo state, Brazil, with different characteristics, were considered: São Paulo, Campinas and Cubatão. The data set of daily PM₁₀ concentration [$\mu\text{g}/\text{m}^3$], relative humidity [%], and ambient temperature [$^{\circ}\text{C}$], were obtained on the Environmental Sanitation Technology Company website [50].

The Brazilian National Health System provides data about the daily hospital admissions due to respiratory diseases (RD). The data set considered in this study, available in [51], comprises the International Classification of Diseases 10 (ICD-10)-J00 to J99. In this work, the database was organized as a daily format and separated by the ICD-10 diagnosis.

According to the Brazilian Institute of Geography and Statistics (IBGE) [52], São Paulo City, the largest city in Brazil, has almost 12 million people (data of 2010) in 1500 km², which is 7398.26 inhabitants per km². The average climate is tropical, about 28 °C in summer and 12 °C in winter [50]. This study considers the period from January 2014 until December 2016. The total number of hospital admissions for respiratory diseases during the studied period, for São Paulo city, was 159,683 occurrences. With regards to the PM₁₀ concentration, only four out of twelve air quality monitoring stations had PM₁₀ data. In addition, only one station presented less than 100 days of lack of data. To deal with this problem, data from another similar station were used to replace them.

Campinas City is the third most populous city of São Paulo State, with a population of approximately 1,1 million people (data of 2010) spread in 795.7 km², a demographic density of 1359.6 inh/km² [52]. The climate is tropical with dry winter and rainy summer with an average of 37 °C during summertime. For this city, the data set considered data from January 2017 to December 2019, comprising 15,464 hospital admissions for respiratory diseases. In this case, two of three air quality monitoring stations presented PM₁₀ data, however, one had no data for 2019. So, the only station with less missing data (145 days lack) was used.

Cubatão has an estimated 118,720 inhabitants with 142.8 km² and 831 inh/km² [52]. In the past, it was one of the most global polluted cities because of its large industrial park and for being surrounded by mountains, which makes the air dispersion hard. In the 1980s, the United Nations considered Cubatão the most polluted city in the world. After that, a government, industries and community effort controlled 98% of the air pollutants level in the city [53]. The current experiments considered the data from January 2017 to December 2019, a total of 802 hospital occurrences. For this city, all three air quality monitoring stations had PM₁₀ available data. However, only the station with more available data was used, with 158 missing days.

A tendency to decrease hospital admissions on the weekends and holidays is a usual situation. For this reason, the day of the week and holidays were considered as two categorical variables [54]. Thus, in addition to the PM₁₀ daily mean concentrations, ambient temperature (T) and relative humidity (RH), the day of the week identifications (1 for Sunday to 7 for Saturday), and a binary flag (h) to recognize if the day is a holiday, were used.

Another important feature is the lag effect of air pollution on human health [10,17,26,55]. A common practice is to consider the effect up to seven days after exposure to air pollution, where lag 0 is the effect on the same day of the exposure, and lag 7 is the effect after seven days of the exposure [54].

Table 1 presents the descriptive statistics for the target (respiratory diseases-RD) and the inputs: PM₁₀ concentration, temperature and relative humidity, for each city. All these variables are differed by average, standard deviation and minimum and maximum values.

Table 1. Descriptive statistics for the variables.

City	Variable	Average	S. Deviation	Min.	Max.
São Paulo	RD	144.0	54.7	9.0	409.0
	PM ₁₀ [µg/m ³]	28.6	14.0	5.0	97.0
	Temperature [°C]	20.7	3.6	9.9	28.9
	Humidity [%]	48.6	16.1	15.0	93.0
Campinas	RD	16.0	6.0	3.0	37.0
	PM ₁₀ [µg/m ³]	21.5	11.3	3.0	84.0
	Temperature [°C]	28.5	3.9	16.6	37.0
	Humidity [%]	42.4	14.4	14.0	90.0
Cubatão	RD	1.0	1.0	0.0	8.0
	PM ₁₀ [µg/m ³]	37.6	17.9	11.0	148.0
	Temperature [°C]	27.1	4.3	16.0	40.3
	Humidity [%]	63.5	16.8	19.0	97.0

Note that the cities have different patterns for the target. São Paulo hospitalizations have a wide dispersion, with 9 to 409 daily hospital admissions. Campinas ranges from 3 to 37, while Cubatão, the smallest studied city, has a maximum of eight hospitalizations. It is necessary to highlight that the databases comprise only data from the public health system, not considering data from health insurance and private units.

The maximum daily PM₁₀ concentration for Cubatão (148 µg/m³) draws attention, because it is almost thrice the WHO 24-hours average limit of 50 µg/m³ (Table 1) [56]. Despite that, the hospital admissions are very low (daily maximum of occurrences) since a significant part of the workers of Cubatão live in São Paulo, which is around 63 km far. The hospital admissions might also depend on the air pollutants dispersion pattern and the local population. São Paulo and Campinas maximum daily PM₁₀ concentrations are lower than Cubatão, but they are also above the WHO limit of 50 µg/m³ (São Paulo-maximum daily of 97 µg/m³; Campinas-maximum daily of 84 µg/m³) [56].

Since the data set described is large with high variability, it may contain multicollinearity or near-linear dependence among the variables. Multicollinearity occurs when two or more inputs (independent variables) are highly correlated affecting the estimate precision. [57]. To evaluate the data set, the Variance Inflation Factor (VIF) shall be used to diagnose the multicollinearity. VIF is calculated by an inflation of the regression coefficient for a independent variable, assessing its correlation to the dependent variables, and modeling the future relation between them. Then, the VIF for each j_{th} factor can be calculated as:

$$VIF_j = \frac{1}{1 - R_j^2}, \quad (8)$$

where R_j^2 is the multiple determination coefficient obtained from regressing each independent variable on the others. If VIF exceeds 5, it is an indicator of multicollinearity [57].

In this work, R Studio (R version 4.1.0 (2021-05-18)—“Camp Pontanezen” Copyright (C) 2021 The R Foundation for Statistical Computing Platform: x86_64 – w64 – mingw32/x64(64 – bit)) was used to calculate VIF. The results are presented in Table 2, showing no multicollinearity between the inputs of each case study.

Table 2. VIF test results for multicollinearity.

VIF	Cubatão	Campinas	São Paulo
PM ₁₀	1.1581	1.5779	1.6365
Relative Humidity	1.9392	2.2771	1.8703
Temperature	1.8825	1.5877	1.2105

In the next section, the proposed models are applied to the presented data, producing a fulfilled analysis of the numerical results obtained.

4. Results and Critical Analysis

The following items describe all models developed to obtain the numerical results in order to evaluate the approach's effectiveness:

- Standard single models: Three versions are developed considering the Standard Models presented in Sections 2.1 and 2.2. The Extreme Learning Machine (ELM), the Echo State Network from Jaeger et al. [29] (ESN J.) and the Echo State Network from Ozturk et al. [44] (ESN O.);
- Regularization Parameter: All standard models are extended, producing three other models through regularization parameter concepts presented in Section 2.3. The ELM with Regularization Parameter (ELM-RP), the ESN J. with Regularization Parameter (ESN J.-RP) and the ESN O. with Regularization Parameter (ESN O.-RP);
- Nonlinear Output Layers: Similarly, three more models are proposed considering the concepts in Section 2.4. The Nonlinear Output Layers strategy is applied to the three single forms creating the ELM with Volterra Filtering Structure (ELM Volt), the ESN J. with Volterra Filtering Structure (ESN J. Volt), and the ESN O. with Volterra Filtering Structure (ESN O. Volt).

The experimental procedure follows the steps summarized in Figure 3:

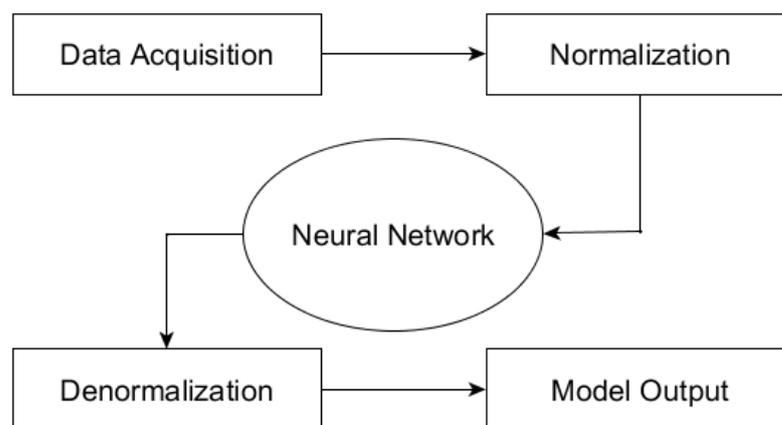


Figure 3. Neural networks appliance steps.

The process begins by collecting the data in the mentioned repositories. Before the insertion of the samples in the neural networks, a normalization procedure is performed due to the limits of the activation function saturation [58]. After the training samples are inserted in the model in order to adjust their free parameters, observing the decrease of the output error. During this process, cross-validation is performed to increase the system generalization capability.

When the training ends, the test samples are inserted in the ANN after the input normalization. The neural response is stored, the normalization is reversed and, finally, the model output is available, which allows the calculation of the models' error. In this work, all models codes were developed in the MATLAB language.

In the training step, the parameters were defined as follows:

- The number of artificial neurons in the hidden layer (or dynamic reservoir) of each model was determined considering a grid search ranging from 3 to 450 neurons;
- The weights were randomly generated in the interval $[-1; +1]$;
- The hyperbolic tangent was addressed as the activation function of the hidden layers;
- The samples were normalized in the interval $[-1; +1]$ before the neural processing;
- The models with RP strategy considered the holdout cross-validation;

- The reservoir designed by Ozturk et al. considered a spectral radius of 0.95 [44];
- The first and the third orders (Equation (7)) of the Volterra filter and the first three principal components of the PCA were considered [48]. These values were defined after empirical tests;
- Before the calculation of the errors, the original domain data was re-scaled.

This work addressed three error metrics to evaluate the solutions quality: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), given by (9)–(11), respectively:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^N (d_n - y_n)^2}, \quad (9)$$

$$\text{MAE} = \frac{1}{N} \sum_{n=1}^N |d_n - y_n|, \quad (10)$$

$$\text{MAPE} = \frac{1}{N} \sum_{n=1}^N \left| \frac{d_n - y_n}{d_n} \right| \times 100, \quad (11)$$

where d_n is the actual value, y_n is the neural model response and N is the total number of samples.

Tables 3–5 present the computational performances achieved by the nine proposed models for each lag, considering each city. The results present the number of neurons (NN) used in the best performance and the error metrics: RMSE, MAE and MAPE. However, as it can be seen in Table 3 for Cubatão, the error metrics MAPE was not considered due to the expressive number of “zeros” for the actual value (d_n). In the tables, the best results obtained for each error metric and the best model are highlighted in purple. Furthermore, the models highlighted in italic bold with stars are the models which obtained statistically similar results to the best one. This statistical test is described below.

A specific result analysis shows that ELM(RP) had the best results for the all calculated metrics for Cubatão in lag 2. Besides, ELM obtained the best results for Campinas, considering the error metrics RMSE and MAPE in lag 3, but for MAE, ELM(RP) achieved the best results in lag 0. For São Paulo, ELM obtained the smallest error values for different lags: RMSE in lag 2 and MAE in lag 1. Finally, ELM(RP) presented the smallest error metric MAPE in lag3.

Note that the best results obtained by the models sometimes were not replicated for all error metrics. This behavior was evident for São Paulo and Campinas, since the best lag and the best model were not always the same. Similar behavior can be observed in [17,59].

The pairwise Wilcoxon test was applied to evaluate if the results are statistically different considering the RMSE with 30 independent simulations [60]. In Tables 3 and 4, the models highlighted in bold with star tag achieved a p -value higher than 0.05, which means that there is no statistical difference between their results and the best one. For this reason, these models can be considered similar, in terms of performance, to the models that obtained the best results. For Campinas, the standard ELM and all ESNs presented equivalent performances, despite the numerical values being contrasts. For Cubatão, ELM and ELM(RP) results were also similar. At long last, for São Paulo the test did not show any statistical similarity among the models.

Figures 4–6 show the boxplot graphic regarding the RMSE values for each city and the lag associated with the best result.

Table 3. Results for Cubatão (Number of neurons-NN, RMSE and MAE for each model and lag).

LAG 0							LAG 1		
Model	NN	RMSE	MAE	NN	RMSE	MAE	NN	RMSE	MAE
ELM	250	1.5630	1.1857	300	1.4760	1.1357			
ELM(RP)	350	1.5330	1.1643	320	1.4808	1.1357			
ELM Volt	350	2.4202	2.0429	450	2.0942	1.7714			
ESN J.	320	1.6058	1.2643	450	1.5789	1.1929			
ESN J.(RP)	450	1.5879	1.2357	450	1.6345	1.2571			
ESN J.Volt	70	2.3815	1.9571	35	1.8323	1.4571			
ESN O.	30	1.6257	1.2143	35	1.4904	1.1429			
ESN O.(RP)	200	1.6797	1.3357	450	1.7587	1.3929			
ESN O.Volt	10	2.7877	2.2286	380	2.7255	2.2429			
LAG 2							LAG 3		
Model	NN	RMSE	MAE	NN	RMSE	MAE	NN	RMSE	MAE
<i>ELM*</i>	420	1.4417	1.1000	350	1.4663	1.1500			
ELM(RP)	320	1.4343	1.0714	380	1.4467	1.1286			
ELM Volt	420	2.0107	1.6929	100	1.9928	1.6571			
ESN J.	300	1.4344	1.1000	380	1.4417	1.1214			
ESN J.(RP)	450	1.5142	1.1786	420	1.4541	1.1000			
ESN J.Volt	30	2.1827	1.7071	35	2.3664	1.9143			
ESN O.	350	1.4760	1.1214	35	1.4880	1.1286			
ESN O.(RP)	420	1.6058	1.2357	170	1.5330	1.2071			
ESN O.Volt	300	2.4900	2.0429	350	2.6227	2.2143			
LAG 4							LAG 5		
Model	NN	RMSE	MAE	NN	RMSE	MAE	NN	RMSE	MAE
ELM	250	1.5071	1.1714	420	1.5353	1.1571			
ELM(RP)	250	1.5024	1.1786	400	1.5306	1.1643			
ELM Volt	280	2.3890	1.9929	380	2.5114	2.1500			
ESN J.	420	1.5142	1.1929	350	1.5561	1.1714			
ESN J.(RP)	450	1.5189	1.1643	350	1.5561	1.2214			
ESN J.Volt	70	2.2960	1.8714	35	2.5746	2.1500			
ESN O.	380	1.6013	1.2786	70	1.5766	1.1571			
ESN O.(RP)	380	1.5561	1.2071	250	1.6191	1.2500			
ESN O.Volt	50	2.3770	2.0500	30	2.4275	1.9643			
LAG 6							LAG 7		
Model	NN	RMSE	MAE	NN	RMSE	MAE	NN	RMSE	MAE
ELM	250	1.4516	1.1500	350	1.5811	1.2286			
ELM(RP)	320	1.5515	1.1929	200	1.5376	1.2000			
ELM Volt	450	2.4640	2.1429	450	2.4928	2.1643			
ESN J.	170	1.6903	1.3429	450	1.5306	1.2214			
ESN J.(RP)	420	1.5584	1.2429	420	1.5834	1.2571			
ESN J.Volt	40	2.3634	2.0429	70	2.2409	1.8786			
ESN O.	40	1.5142	1.1714	35	1.5811	1.2429			
ESN O.(RP)	250	1.6410	1.3357	200	1.5969	1.3071			
ESN O.Volt	300	2.8322	2.2143	50	2.6390	2.1071			

Table 4. Results for Campinas (Number of neurons-NN, RMSE, MAE and MAPE for each model and lag).

LAG 0					LAG 1			
Model	NN	RMSE	MAE	MAPE %	NN	RMSE	MAE	MAPE %
ELM	25	6.9017	5.6479	40.2496	3	5.5462	4.4507	36.9895
ELM(RP)	3	5.1094	3.9648	32.7044	25	7.0751	5.7324	40.8537
ELM Volt	25	7.0206	5.2676	37.2186	10	5.3910	4.1338	34.4946
ESN J.	35	6.9394	5.6620	40.0763	50	6.6619	5.4085	41.5375
ESN J.(RP)	3	6.4306	5.0563	50.5484	3	6.4731	5.0845	51.6402
ESN J.Volt	380	6.3540	4.8803	46.8265	3	5.2393	3.9577	33.7701
ESN O.	15	5.8713	4.6127	39.4574	30	6.4878	5.2465	40.5271
ESN O.(RP)	3	6.2473	4.9014	49.1485	7	6.6327	5.2324	52.7208
ESN O.Volt	3	5.6438	4.3873	33.9491	3	5.8743	4.6127	34.6702
LAG 2					LAG 3			
Model	NN	RMSE	MAE	MAPE %	NN	RMSE	MAE	MAPE %
ELM	15	6.4464	5.1972	38.4853	3	5.0644	4.0282	31.9037
ELM(RP)	3	5.4721	4.3662	33.0808	25	6.6072	5.1761	39.8549
ELM Volt	25	5.9517	4.4507	38.5412	170	6.2020	4.6268	41.3376
<i>ESN J.*</i>	30	6.4114	5.0915	41.0724	70	6.1260	4.7113	38.2705
<i>ESN J.(RP)*</i>	3	6.2258	4.7887	48.3773	10	6.2196	4.9296	49.2648
<i>ESN J.Volt*</i>	3	5.7684	4.4859	35.2526	30	5.7101	4.2817	38.4659
<i>ESN O.*</i>	25	6.2557	4.9085	39.6942	100	6.2905	4.8310	37.4986
<i>ESN O.(RP)*</i>	10	6.0630	4.6761	46.9116	5	6.3207	4.9577	49.5265
<i>ESN O.Volt*</i>	7	5.7648	4.3592	40.2773	450	6.1633	4.8732	42.5624
LAG 4					LAG 5			
Model	NN	RMSE	MAE	MAPE %	NN	RMSE	MAE	MAPE %
ELM	3	5.7403	4.4648	32.2381	3	5.2928	4.0845	33.6603
ELM(RP)	20	6.6003	5.0986	37.0427	3	5.3200	4.1056	34.3353
ELM Volt	25	5.7885	4.4085	34.7377	30	5.9511	4.7394	37.2746
ESN J.	70	6.2054	4.7746	36.7789	35	6.1070	4.7042	36.9522
ESN J.(RP)	30	6.3184	4.9859	48.8737	3	6.1254	4.7183	46.9130
ESN J.Volt	35	5.2682	4.1056	33.8667	3	5.8934	4.6479	35.1359
ESN O.	120	6.2776	4.8310	36.6999	200	6.3745	4.9859	38.5786
ESN O.(RP)	7	5.9935	4.6831	46.2167	10	6.2377	4.8239	47.9457
ESN O.Volt	3	5.9570	4.3028	36.7693	3	5.4521	4.3732	34.0759
LAG 6					LAG 7			
Model	NN	RMSE	MAE	MAPE %	NN	RMSE	MAE	MAPE %
ELM	3	5.3068	4.1972	36.8548	35	7.2452	5.6761	42.0235
ELM(RP)	3	5.2474	4.0704	36.8444	35	7.2384	5.6761	42.0908
ELM Volt	25	5.3253	4.2465	35.6237	70	5.1273	4.0930	38.2203
ESN J.	30	6.2360	4.9930	39.2345	50	6.6961	5.1620	41.5200
ESN J.(RP)	5	5.9741	4.6901	44.9676	3	5.8928	4.6549	45.2283
ESN J.Volt	3	5.7873	4.4930	35.0669	3	6.0082	4.8169	34.6062
ESN O.	35	6.3987	5.0563	39.7110	50	6.4579	4.8732	39.7276
ESN O.(RP)	5	5.9487	4.5000	44.5221	5	5.9871	4.6620	45.6153
ESN O.Volt	3	5.6519	4.4014	35.8808	3	5.6687	4.2746	37.4182

Table 5. Results for São Paulo (Number of neurons-NN, RMSE, MAE and MAPE for each model and lag).

LAG 0					LAG 1			
Model	NN	RMSE	MAE	MAPE %	NN	RMSE	MAE	MAPE %
ELM	25	61.1156	51.141	43.4189	3	39.7697	30.7821	36.1940
ELM(RP)	15	55.2425	46.4231	41.1963	20	59.1541	48.3205	40.3268
ELM Volt	10	60.8374	48.2179	48.3650	20	72.4500	58.7179	57.4171
ESN J.	420	62.4953	51.3910	42.4775	100	62.1230	50.4167	42.4388
ESN J.(RP)	450	67.7300	55.0449	70.2352	450	67.6519	55.0769	69.8433
ESN J.Volt	400	66.9744	54.0705	64.0419	3	51.2727	40.5833	43.2995
ESN O.	50	58.7160	48.3141	39.7660	50	58.1013	45.9679	40.5099
ESN O.(RP)	380	69.2728	56.6154	71.5779	350	68.8875	55.8526	71.2675
ESN O.Volt	3	57.8689	46.5064	45.7923	3	59.6593	45.7628	50.0376
LAG 2					LAG 3			
Model	NN	RMSE	MAE	MAPE %	NN	RMSE	MAE	MAPE %
ELM	3	39.3745	31.2628	36.9582	25	59.9251	48.2308	45.2785
ELM(RP)	20	60.6640	48.2115	41.4840	3	43.9040	35.1026	35.9396
ELM Volt	25	83.6339	69.4423	68.6818	20	71.2603	55.4679	57.0193
ESN J.	70	60.7151	49.1154	42.8314	100	64.7131	52.4423	48.2804
ESN J.(RP)	450	65.7486	53.6795	68.8934	400	65.3756	53.0064	69.0983
ESN J.Volt	3	59.2428	43.6090	46.8035	3	58.6061	44.1795	43.8522
ESN O.	35	58.6986	47.4103	40.7649	15	50.4066	39.3910	41.1334
ESN O.(RP)	420	68.1032	55.3974	70.3439	400	67.3167	54.4936	70.1148
ESN O.Volt	3	55.5113	43.2628	45.1094	5	58.2715	45.4872	46.8127
LAG 4					LAG 5			
Model	NN	RMSE	MAE	MAPE %	NN	RMSE	MAE	MAPE %
ELM	15	53.6898	41.4167	43.9006	3	46.4592	37.1474	39.0748
ELM(RP)	3	45.6739	35.9231	42.4408	3	43.6788	33.3077	38.8188
ELM Volt	7	64.8803	49.4487	56.4511	7	49.8634	39.3974	47.2036
ESN J.	450	64.3933	51.3782	47.0148	100	63.9561	51.8141	47.3902
ESN J.(RP)	450	65.9862	52.4808	69.1144	380	66.2419	52.6603	68.9928
ESN J.Volt	380	72.9358	59.3526	64.7092	7	68.9339	55.0513	56.2654
ESN O.	15	55.3579	45.3654	44.0190	25	58.9842	45.2051	44.8766
ESN O.(RP)	380	69.3555	55.5513	72.2436	380	68.0724	54.3590	70.6102
ESN O.Volt	3	50.5135	40.5833	40.8100	3	53.3191	41.0000	48.6203
LAG 6					LAG 7			
Model	NN	RMSE	MAE	MAPE %	NN	RMSE	MAE	MAPE %
ELM	20	54.3250	43.9744	43.4307	25	55.3315	43.9744	39.5735
ELM(RP)	3	47.3653	36.9551	39.9839	3	44.0574	35.1923	37.5251
ELM Volt	20	64.5582	46.3654	49.4889	10	75.8567	63.5641	63.1259
ESN J.	150	59.9074	49.0641	45.4045	70	58.9236	47.3141	40.8857
ESN J.(RP)	380	66.0817	52.5705	68.6722	380	65.4964	52.9615	68.4640
ESN J.Volt	3	60.9871	45.6859	47.4705	3	62.1758	47.5641	45.0664
ESN O.	20	54.1340	43.2756	43.7886	35	51.7540	42.4103	39.8727
ESN O.(RP)	420	68.5719	54.7949	70.9594	320	67.4013	54.7564	70.0246
ESN O.Volt	3	57.0886	43.9487	46.1440	3	53.9567	43.3974	44.4320

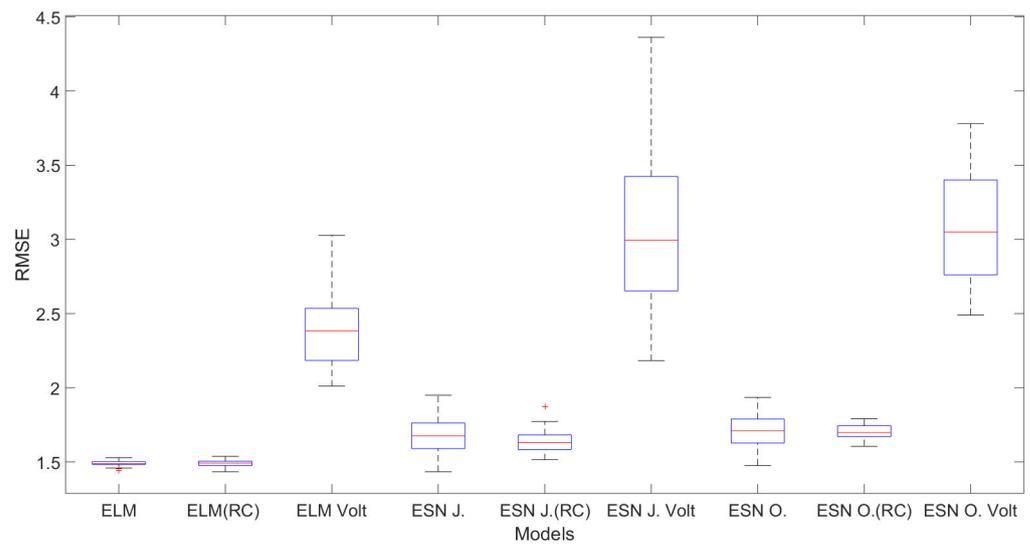


Figure 4. Boxplot graphic regarding the RMSE values for Cubatão-Lag 2.

Considering Cubatão, observe that the smallest dispersion was obtained by ELM (RP) model, which also presented the smallest average value, corroborating the observation from Table 3. The inclusion of the Volterra filter increases the dispersion and the average values for all standard models, representing a significant degree of deterioration in performance.

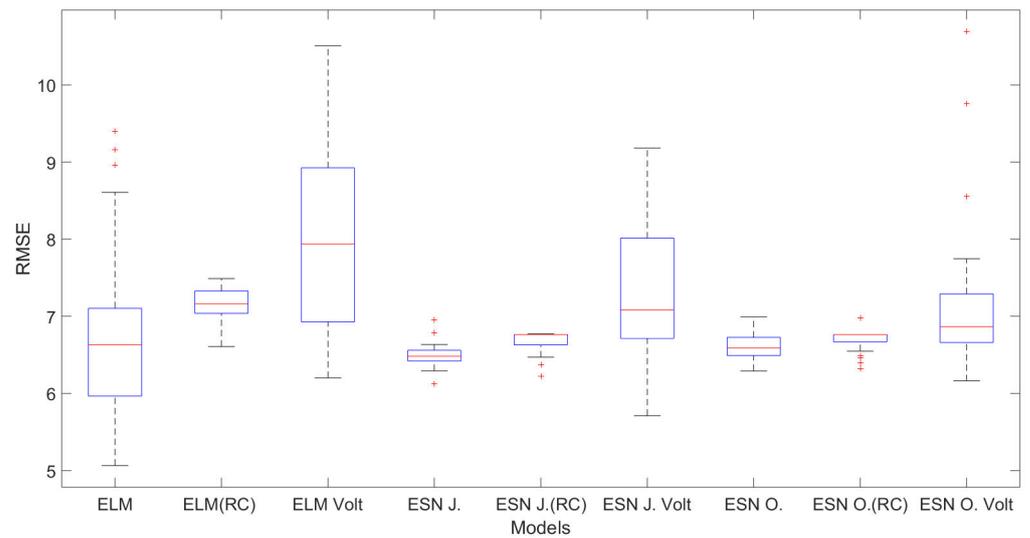


Figure 5. Boxplot graphic regarding the RMSE values for Campinas-Lag 3.

In Campinas' case, only the ELM performances will be considered in this specific analysis since all ESN obtained similar results according to the Wilcoxon test. However, Figure 5 illustrates all models to avoid any curiosity. The RP inclusion decreases the dispersion, while the Volterra filter showed an opposite behavior. Despite that, the best performance in terms of best results regarding 30 simulations was favorable to the use of Volterra filter instead of the RP (note the bottom value in the boxplot). Since the generation of the neurons' weights were random, the algorithms must run at least 30 times, and this fact directly implied a long tail for the boxplots, as can be seen in the Volterra models. Moreover, the best result obtained by ELM does not mean the best performance in terms of dispersion.

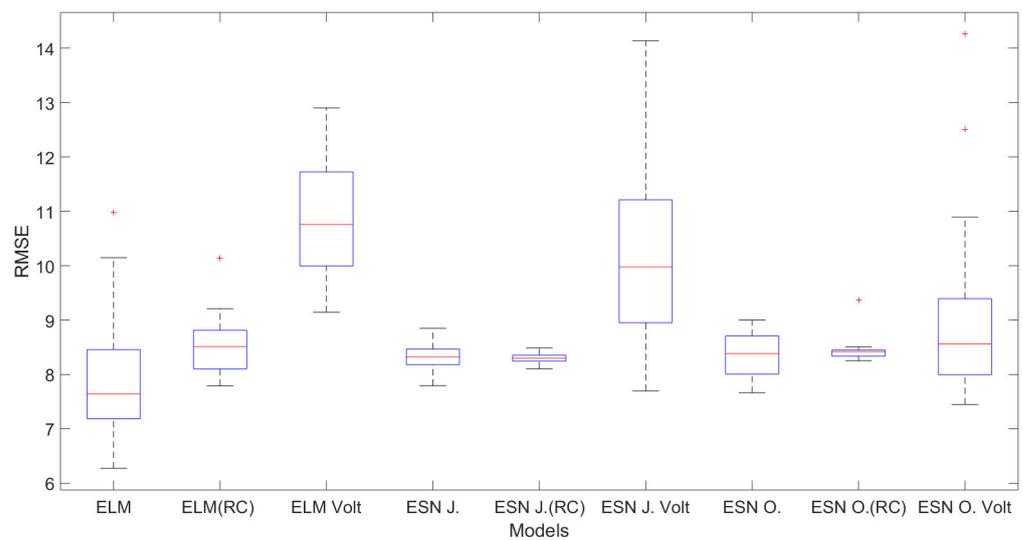


Figure 6. Boxplot graphic regarding the RMSE values for São Paulo-Lag 2.

For São Paulo, the general behavior of the standard models was similar to Campinas. The standard ELM achieved better general errors, even when considered the median value. The inclusion of the RP reduced the dispersion, but it decreased the probability of obtaining better results for the error metrics. On the other hand, the Volterra filter showed a worse performance in terms of dispersion. However, despite the ELM best results for error metric, the model presented a bad dispersion, including an outlier.

Table 6 presents a ranking of best error metric results considering all neural models in ascending order of development. Note that the draws regarding the winners mean that there was no statistical difference between the models. The last column represents the final ranking considering the three cities’ results.

Table 6. Ranking of the models’ performance.

Model	Cubatão (lag 2)		Campinas(lag 3)			São Paulo (lag 2)			Mean	Rank
	RMSE	MAE	RMSE	MAE	MAPE %	RMSE	MAE	MAPE %		
ELM	1	1	1	1	1	1	1	1	1	1st
ELM(RP)	1	1	9	9	8	5	5	3	5.1	8th
ELM Volt	7	7	8	8	9	9	9	7	8.0	9th
ESN J.	3	3	1	1	1	6	6	4	3.12	3rd
ESN J.(RP)	6	6	1	1	1	7	7	8	4.6	6th
ESN J.Volt	8	8	1	1	1	4	3	6	4.0	5th
ESN O.	4	4	1	1	1	3	4	2	2.5	2nd
ESN O.(RP)	5	5	1	1	1	8	8	9	4.8	7th
ESN O.Volt	9	9	1	1	1	2	2	5	3.8	4th

The standard ELM was the best estimator in all cases as regards the error metric results, but for Cubatão, the results obtained by ELM(RP) were the same. The second and third positions show ESN O. and ESN J. models, respectively. However, despite the main contribution of RP is to increase the models’ generalization capability, its use reduced the dispersion of the results, i.e., the models’ predictability increased, except for Cubatão. Moreover, the ELM(RP) ranking position was deteriorated by the Campinas results, since all ESNs presented the same statistical performance. Dismissing these aspects, the model could be the second best.

Although the inclusion of the Volterra filter did not improve the performances, the idea of its application was to capture nonlinear patterns among the signals from the hidden layer. Despite the literature presents good performances for this method in correlated tasks [33], its use is not recommended in this case. Similarly, the inclusion of the reservoir designed by Jaeger or Ozturk et al. is not adequate to the problem.

Regarding the number of neurons in the hidden layers (dynamic reservoir), one can see miscellaneous neurons, with a high degree of variation. For Cubatão, the pattern noted was the models used hundreds of neurons in most cases. Interestingly, ESN J. and ESN O. often addressed up to 70 neurons. Moreover, it can be seen in Campinas' case, that the RP models used up to 35 neurons in all cases. Considering São Paulo, the ELM versions tended to use less than 25 neurons, similar to ESN J. Volt, ESN O. and ESN O. Volt. The others models addressed hundreds of neurons. This is a strong indication that a sweep in the neuron amount is needed because a clear pattern regarding this parameter was not found. Even considering the results of the models that presented a p -value large than 0.05, the number of neurons was variable.

In summary, the unorganized machines are particular cases of classic neural models, which the hidden weights are not adjusted. On one hand, the user may lose part of the approximation capability due to this characteristic; on the other hand, there are gains in terms of training effort and stability for the output values during the training, avoiding discrepancies. An important aspect is that these methodologies can be outperformed, depending on the problem. Regarding the use of RP or Volterra filter, the literature indicates that these strategies may increase the mapping capability of the neural models. However, this work showed that in specific cases these approaches did not present efficiency.

Figures 7–9 present the best evolution of the output response in comparison to the actual values.

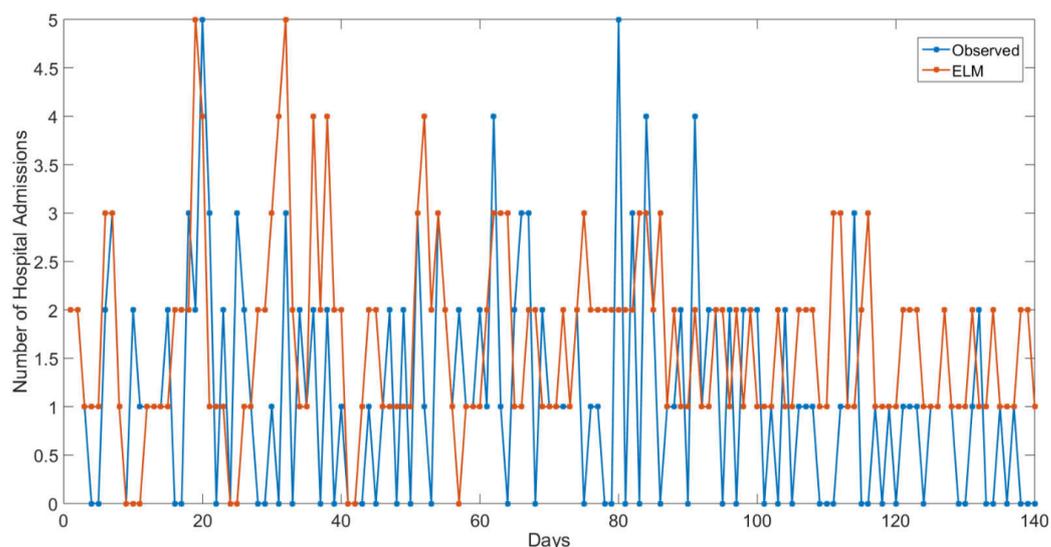


Figure 7. The number of hospital admissions by day of the test set for ELM lag 2-Cubatão (observed versus estimated values).

Figure 7 shows that the prediction task seems to be more difficult when the output has a small range and many “zero” observations. In this case, as the overestimation was small, given that the observed values are zero, it did not interfere in hospital management. Otherwise, in Figure 8, since there were no “zero” observations, the ELM estimations could be considered a suitable performance, except in abrupt cases.

Finally, in Figure 9, ELM reached the smallest RMSE, but comparing with the observed data, it was more difficult to predict the abrupt decrease of hospital admissions occurred around day 70. On the other hand, the ELM(RP) could follow this tendency, but it over and underestimated the number of hospital admissions in many cases. These behaviors are directly related to the number of neurons used by each model since a reduction in this number limited the model approximation capability.

Regarding the best error metric to be used, RMSE seems to be a good strategy, since the error metric was reduced during the neural models training (adjustment) [17,18,61].

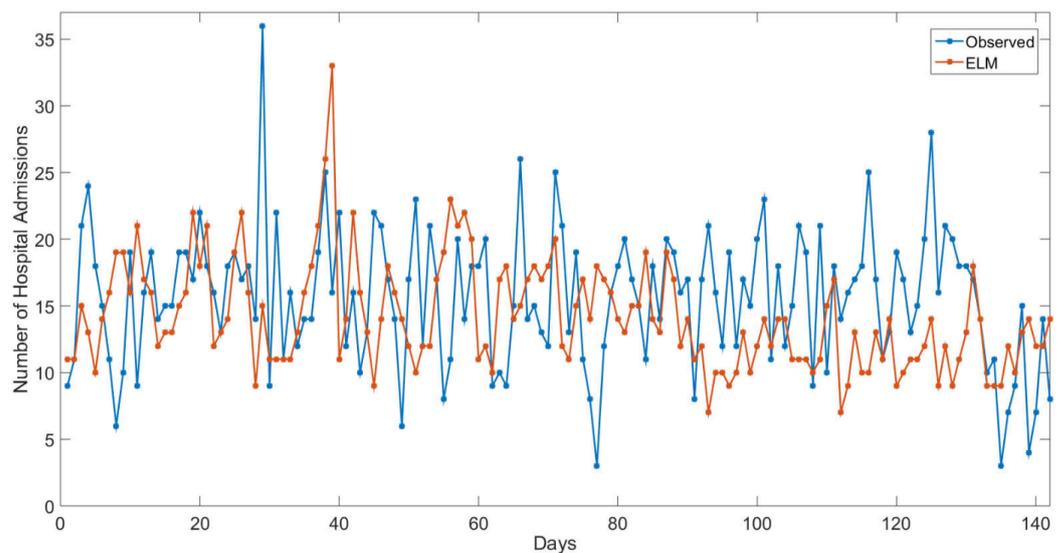


Figure 8. The number of hospital admissions by day of the test set for ELM lag 3-Campinas (observed versus estimated values).

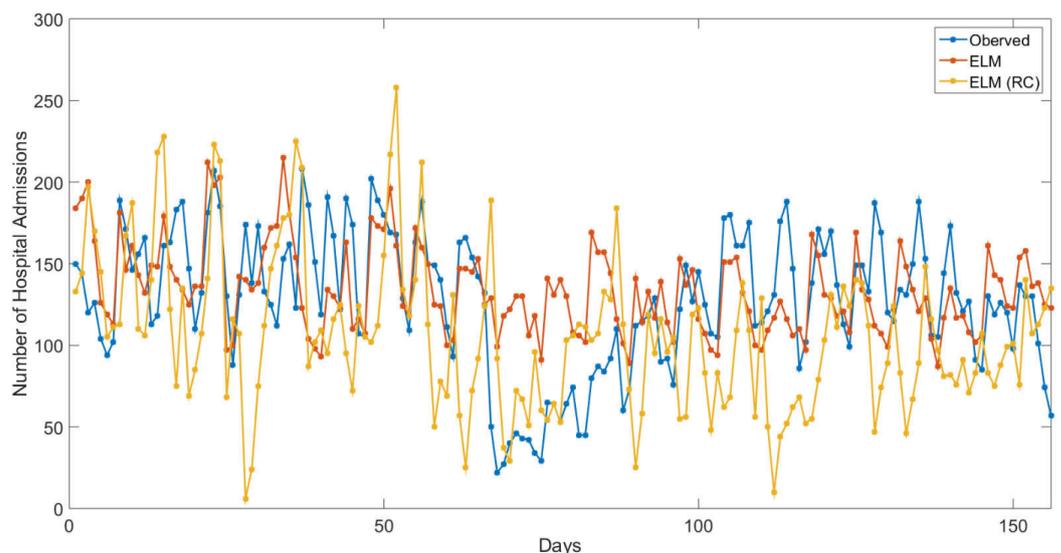


Figure 9. The number of hospital admissions by day of the test set for ELM lag 2-São Paulo (observed versus estimated values).

Table 7 presents a summary of some notable studies showing the association between air pollutants concentration, morbidity (Hospital Admissions or Hospital Emergency) and mortality. This brief description relates the authors, geographic area, considered inputs and predicted variables, the applied methods, metrics, time base, and the best MAPE and RMSE observed for each study. Although these studies present suitable estimations and relevant contributions, they proposed different models, and applied to diverse worldwide places, using specific inputs to predict health effects. For this reason, a comparative analysis of these studies' performances is unfair, as Katri and Tamil [62] previously observed. However, some important aspects can be highlighted.

Two studies [62,63] did not use MAPE or RMSE as error metrics. Khatri and Tamil [62] aimed to compare the performance for peak and non-peak class prediction. The authors used percentage difference in this study and applied MLP, without any consideration about other methods' performance. Shakerkhatibi et al. [64] used other metrics (Delong Method) to compare the predictions using MLP and Conditional Logistic Regression.

Table 7. Summary of studies presenting air pollutant’s associations with morbidity and mortality using ANN.

Authors (Year)	Geographic Area of Study	Inputs	Predicted Variable	Methods	Used Metrics	Time Base	Best MAPE	Best RMSE
Kassomenos et al. (2011) [24]	Athens	T, RH, WD, SO ₂ , black smoke CO, NO ₂ , NO, O ₃	HA for Cardiorespiratory diseases	MLP, GLM	RMSE	daily	NA	0.8950
Moustris et al. (2012) [63]	Athens	T, RH, WS, solar radiation SO ₂ , PM ₁₀ , CO, O ₃ , NO ₂ (age subgroups 0–4 years, 5–14 years, 0–14 years)	HA for Asthma	MLP (TLRN)	MBE, RMSE, R ² , IA	daily	NA	3.2
Cengiz and Terzi (2012) [65]	Afyon, Turkey	SO ₂ , PM ₁₀	HA and symptoms (cough, exertional, dyspnea, expectoration) for COPD	MLP, RBF, GLM, GAM	RMSE e MAPE	weekly	4.54	2.38
Shakerkhatibi et al. (2015) [64]	Tabriz, Iran	T, RH, NO, NO ₂ , NO _x , SO ₂ , CO, PM ₁₀ , O ₃ (age and gender subgroups)	HA for respiratory and cardiovascular diseases	MLP, CLR	AUC, sensitivity, Specificity and Accuracy (%)	daily	NA	NA
Khatri and Tamil (2017) [62]	Dallas County, Texas, USA	T, RH, WS, CO, O ₃ , SO ₂ , NO ₂ , PM _{2.5}	HE for respiratory diseases	MLP	% difference	daily	NA	NA
Tadano et al. (2016) [26]	Campinas city, São Paulo state, Brazil	T, RH, PM ₁₀	HA for respiratory diseases	MLP, ESN, ELM	MSE/MAPE	daily	31.2	5.98
Polezer et al. (2018) [10]	Curitiba, Paraná, Brazil	T, RH, PM _{2.5}	HA for respiratory diseases	MLP, ESN, ELM	MSE/MAPE	daily	29.87	7.37
Araujo et al. (2020) [17]	Campinas and São Paulo cities, Brazil	T, RH, PM ₁₀	HA for respiratory diseases	MLP, GLM, ELM, ESN, RBF, Ensemble	MSE, MAE, MAPE	daily	24.87	3.04
Zhou, Li and Wang (2018) [66]	Hangzhou, Southern part of the Yangtze River Delta, China	T, PM ₁₀ , PM _{2.5} , NO ₂ , SO ₂	Respiratory disease cases	MLP, GAM	AIC, MSE	daily	NA	2.17
Kachba et al. (2020) [18]	São Paulo city, Brazil	CO, NO _x , O ₃ , SO ₂ , PM	HA and mortality for respiratory diseases	MLP, ELM, ESN	MSE, MAE, MAPE	monthly	34.53	160.26

WD-Wind Direction; WS-Wind Speed; SO₂-Sulphur Dioxide; CO-Carbon Monoxide; NO₂-Nitrogen Dioxide; NO-Nitrogen Monoxide; O₃-Ozone; NO_x-Nitrogen Oxides; HE-Hospital Emergency; GAM-Generalized Additive Models; TLRN-Time Lagged Recurrent Networks; RBF-Radial Basis Function Network; CLR-Conditional Logistic Regression Modeling; MBE- Mean Bias Error; IA-Index of Agreement; AUC-area under curve.

Considering the variety of applied methods (Table 7), and emphasizing the use of MLP, the performance comparison between ANN and regression models has proved the ANN superior performance. Inspired by these all aspects, the paper's authors believe that this present work, which explores the ELM and ESN models with variations from the RP and the Volterra filter to estimate hospital admissions due to respiratory diseases caused by air pollutants concentration, is a relevant contribution. However, given the harmful effects of PM on human health, and comparing the considered input variables used in the other studies, this work has some limitations, such as the use of only one air pollutant (PM₁₀) and the lack of comparison with a statistical regression modeling.

5. Conclusions

This work predicted the hospital admissions due to respiratory diseases caused by the particulate matter PM₁₀ concentrations using the extreme learning machines (ELM) and the echo state networks (ESN) in the standard forms and applying the variations from the regularization parameter (RP) and the Volterra filter. The estimates considered daily PM₁₀ concentration, relative humidity, ambient temperature as inputs and predicted the daily hospital admissions for respiratory diseases.

Numerical results indicated the superior performance of the standard models, pointing to ELM as the best predictor for most scenarios. However, regarding Campinas city and the RMSE error metric, a statistical test demonstrated that ESN models were statistically similar when compared to the best one. Besides, a graphic analysis showed that the models with the inclusion of RP strategy presented a reduced dispersion, considering the abrupt variations in hospital admissions, while the Volterra filter showed an opposite behavior, indicating that its application was not suitable for this specific problem. Finally, completing the critical analysis, a ranking of performances classified the models regarding the error metrics for each city. This ranking rewarded the models with statistical similarity rather than models with good dispersion, highlighting the standard models in the first positions.

The application of Unorganized Machines to three different cities was essential to evaluate their good performance in predicting air pollution impacts on human health. An additional graphic analysis of the output response in comparison to the actual values, for the best models, evidenced the good performance of the neural networks to estimate the hospital admissions. This contribution may help governmental bodies and policymakers on the management of hospital planning, mainly during air pollution unfavorable climate periods. Moreover, the good performance of the models confirms the link between all input variables and the output values, verifying that the particulate matter, temperature and relative humidity are fundamental to obtain a good estimation.

A limitation of this study is the lack of large data sets that could bring more uniform performances between the studied cities. As a consequence of the lack of monitoring data, other pollutants variations such as PM_{2.5} cannot be studied.

Considering the continental dimension of Brazil and the characteristics of the different region's climates, it would be paramount to study all regions (states), a hard task due to the lack of monitoring all over the country. Further works shall consider hybrid modeling or ensembles, the use of deseasonalization techniques, and the appliance of other artificial neural networks. Since the ELM is admittedly susceptible to the neurons number changes in the hidden layer and the ESN model is considered robust in this regard, a comparison study should be conducted pointing to the training time required between these models.

Author Contributions: Conceptualization, Y.d.S.T., E.T.B., L.C., and H.V.S.; methodology, H.V.S. and Y.d.S.T.; software, H.V.S., T.S.P., E.P., and T.A.A.; formal analysis, Y.d.S.T., E.T.B., and L.C.; investigation, H.V.S., E.P., T.A.A., and Y.d.S.T.; resources, H.V.S.; data curation, E.T.B. and L.C.; writing—original draft preparation, E.T.B., L.C., and H.V.S.; writing—review and editing, Y.d.S.T. and C.M.L.U.; visualization, H.V.S. and T.A.A.; supervision, Y.d.S.T. and H.V.S.; project administration, H.V.S.; funding acquisition, H.V.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Council for Scientific and Technological Development (CNPq), grant number 405580/2018-5, and the APC was funded by DIRPPG/UTFPR/PG.

Acknowledgments: The authors thank the Brazilian agencies Coordination for the Improvement of Higher Education Personnel (CAPES)-Financing Code 001, Brazilian National Council for Scientific and Technological Development (CNPq), processes number 40558/2018-5, 315298/2020-0, and Araucaria Foundation, process number 51497, and Federal University of Technology-Parana (UTFPR) for their financial support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. WHO-World Health Organization. *Ambient Air Pollution: Health Impacts*; WHO: Geneva, Switzerland, 2018.
2. Lelieveld, J.; Evans, J.S.; Fnais, M.; Giannadaki, D.; Pozzer, A. The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature* **2015**, *525*, 367. [[CrossRef](#)] [[PubMed](#)]
3. Manisalidis, I.; Stavropoulou, E.; Stavropoulos, A.; Bezirtzoglou, E. Environmental and health impacts of air pollution: A review. *Front. Public Health* **2020**, *8*, 14. [[CrossRef](#)] [[PubMed](#)]
4. Li, X.; Liu, X. Effects of PM_{2.5} on chronic airway diseases: A review of research progress. *Atmosphere* **2021**, *12*, 1068. [[CrossRef](#)]
5. Ab Manan, N.; Aizuddin, A.N.; Hod, R. Effect of air pollution and hospital admission: A systematic review. *Ann. Glob. Health* **2018**, *84*, 670. [[CrossRef](#)] [[PubMed](#)]
6. Grigorieva, E.; Lukyanets, A. Combined effect of hot weather and outdoor air pollution on respiratory health: Literature review. *Atmosphere* **2021**, *12*, 790. [[CrossRef](#)]
7. Morrissey, K.; Chung, I.; Morse, A.; Parthasarath, S.; Roebuck, M.M.; Tan, M.P.; Wood, A.; Wong, P.F.; Forstick, S.P. The effects of air quality on hospital admissions for chronic respiratory diseases in Petaling Jaya, Malaysia, 2013–2015. *Atmosphere* **2021**, *12*, 1060. [[CrossRef](#)]
8. Yitshak-Sade, M.; Nethery, R.; Schwartz, J.D.; Mealli, F.; Dominici, F.; Di, Q.; Awad, Y.A.; Ifergane, G.; Zanobetti, A. PM_{2.5} and hospital admissions among Medicare enrollees with chronic debilitating brain disorders. *Sci. Total Environ.* **2021**, *755*, 142524. [[CrossRef](#)]
9. Anderson, J.O.; Thundiyil, J.G.; Stolbach, A. Clearing the air: A review of the effects of particulate matter air pollution on human health. *J. Med. Toxicol.* **2012**, *8*, 166–175. [[CrossRef](#)]
10. 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. Assessing the impact of PM_{2.5} on respiratory disease using artificial neural networks. *Environ. Pollut.* **2018**, *235*, 394–403. [[CrossRef](#)]
11. Ardiles, L.G.; Tadano, Y.S.; Costa, S.; Urbina, V.; Capucim, M.N.; da Silva, I.; Braga, A.; Martins, J.A.; Martins, L.D. Negative binomial regression model for analysis of the relationship between hospitalization and air pollution. *Atmos. Pollut. Res.* **2018**, *9*, 333–341. [[CrossRef](#)]
12. McCullagh, P.; Nelder, J.A. *Generalized Linear Models*; Routledge: London, UK, 2019.
13. Belotti, J.T.; Castanho, D.S.; Araujo, L.N.; da Silva, L.V.; Alves, T.A.; Tadano, Y.S.; Stevan, S.L., Jr.; Correa, F.C.; Siqueira, H.V. Air Pollution Epidemiology: A Simplified Generalized Linear Model Approach Optimized by Bio-Inspired Metaheuristics. *Environ. Res.* **2020**, *191*, 110106. [[CrossRef](#)] [[PubMed](#)]
14. Cromar, K.; Galdson, L.; Palomera, M.J.; Perlmutter, L. Development of a health-based index to identify the association between air pollution and health effects in Mexico City. *Atmosphere* **2021**, *12*, 372. [[CrossRef](#)]
15. Ravindra, K.; Rattan, P.; Mor, S.; Aggarwal, A.N. Generalized additive models: Building evidence of air pollution, climate change and human health. *Environ. Int.* **2019**, *132*, 104987. [[CrossRef](#)] [[PubMed](#)]
16. Zhou, H.; Geng, H.; Dong, C.; Bai, T. The short-term harvesting effects of ambient particulate matter on mortality in Taiyuan elderly residents: A time-series analysis with a generalized additive distributed lag model. *Ecotoxicol. Environ. Saf.* **2021**, *207*, 111235. [[CrossRef](#)]
17. Araujo, L.N.; Belotti, J.T.; Antonini Alves, T.; de Souza Tadano, Y.; Siqueira, H. Ensemble method based on Artificial Neural Networks to estimate air pollution health risks. *Environ. Model. Softw.* **2020**, *123*, 104567. [[CrossRef](#)]
18. Kachba, Y.; Chiroli, D.M.d.G.; Belotti, J.T.; Antonini Alves, T.; de Souza Tadano, Y.; Siqueira, H. Artificial Neural Networks to Estimate the Influence of Vehicular Emission Variables on Morbidity and Mortality in the Largest Metropolis in South America. *Sustainability* **2020**, *12*, 2621. [[CrossRef](#)]
19. Cabaneros, S.M.; Calautit, J.K.; Hughes, B.R. A review of artificial neural network models for ambient air pollution prediction. *Environ. Model. Softw.* **2019**, *119*, 285–304. [[CrossRef](#)]
20. de Mattos Neto, P.S.; Madeiro, F.; Ferreira, T.A.; Cavalcanti, G.D. Hybrid intelligent system for air quality forecasting using phase adjustment. *Eng. Appl. Artif. Intell.* **2014**, *32*, 185–191. [[CrossRef](#)]
21. Feng, R.; Zheng, H.J.; Gao, H.; Zhang, A.R.; Huang, C.; Zhang, J.X.; Luo, K.; Fan, J.R. Recurrent Neural Network and random forest for analysis and accurate forecast of atmospheric pollutants: A case study in Hangzhou, China. *J. Clean. Prod.* **2019**, *231*, 1005–1015. [[CrossRef](#)]

22. Neto, P.S.D.M.; Firmino, P.R.A.; Siqueira, H.; Tadano, Y.D.S.; Alves, T.A.; De Oliveira, J.F.L.; Marinho, M.H.D.N.; Madeiro, F. Neural-Based Ensembles for Particulate Matter Forecasting. *IEEE Access*. **2021**, *9*, 14470–14490. [[CrossRef](#)]
23. Wang, Q.; Liu, Y.; Pan, X. Atmosphere pollutants and mortality rate of respiratory diseases in Beijing. *Sci. Total Environ.* **2008**, *391*, 143–148. [[CrossRef](#)]
24. Kassomenos, P.; Petrakis, M.; Sarigiannis, D.; Gotti, A.; Karakitsios, S. 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* **2011**, *4*, 263–272. [[CrossRef](#)]
25. Sundaram, N.M.; Sivanandam, S.; Subha, R. Elman neural network mortality predictor for prediction of mortality due to pollution. *Int. J. Appl. Eng. Res* **2016**, *11*, 1835–1840.
26. Tadano, Y.S.; Siqueira, H.V.; Antonini Alves, T. Unorganized machines to predict hospital admissions for respiratory diseases. In Proceedings of the IEEE Latin American Conference on Computational Intelligence (LA-CCI), Cartagena, Colombia, 2–4 November 2016; pp. 1–6.
27. Boccato, L.; Soares, E.S.; Fernandes, M.M.L.P.; Soriano, D.C.; Attux, R. Unorganized Machines: From Turing’s Ideas to Modern Connectionist Approaches. *Int. J. Nat. Comput. Res. (IJNCR)* **2011**, *2*, 1–16. [[CrossRef](#)]
28. Huang, G.; Huang, G.B.; Song, S.; You, K. Trends in extreme learning machines: A review. *Neural Netw.* **2015**, *61*, 32–48. [[CrossRef](#)] [[PubMed](#)]
29. Jaeger, H. The “echo state” approach to analysing and training recurrent neural networks—with an erratum note. *Bonn, Ger. Ger. Natl. Res. Cent. Inf. Technol. GMD Tech. Rep.* **2001**, *148*, 13.
30. Jaeger, H. *Short term memory in Echo State Networks*; Technical Report; Fraunhofer Institute for Autonomous Intelligent Systems: Sankt Augustin, Germany, 2001.
31. Siqueira, H.V.; Boccato, L.; Attux, R.; Lyra Filho, C. Echo state networks in seasonal streamflow series prediction. *Learn. Nonlinear Model.* **2012**, *10*, 181–191. [[CrossRef](#)]
32. Huang, G.B.; Zhou, H.; Ding, X.; Zhang, R. Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst. Man, Cybern. Part B Cybern.* **2011**, *42*, 513–529. [[CrossRef](#)]
33. Boccato, L.; Lopes, A.; Attux, R.; Von Zuben, F.J. An extended echo state network using Volterra filtering and principal component analysis. *Neural Networks Off. J. Int. Neural Netw. Soc.* **2012**, *32*, 292–302. [[CrossRef](#)]
34. Butcher, J.; Verstraeten, D.; Schrauwen, B.; Day, C.; Haycock, P. Extending reservoir computing with random static projections: A hybrid between extreme learning and RC. In Proceedings of the 18th European Symposium on Artificial Neural Networks, Bruges, Belgium, 28–30 April 2010.
35. Yildiz, I.B.; Jaeger, H.; Kiebel, S. Re-visiting the echo state property. *Neural Netw.* **2012**, *35*, 1–9. [[CrossRef](#)]
36. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [[CrossRef](#)]
37. Huang, G.; Chen, L.; Siew, C. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans. Neural Netw.* **2006**, *17*, 879–92. [[CrossRef](#)]
38. Cao, J.; Lin, Z.; Huang, G.B.; Liu, N. Voting based extreme learning machine. *Inf. Sci.* **2012**, *185*, 66–77. [[CrossRef](#)]
39. Siqueira, H.; Luna, I. Performance comparison of feedforward neural networks applied to streamflow series forecasting. *Math. Eng. Sci. Aerosp. (MESA)* **2019**, *10*, 41–53.
40. Bartlett, P. The Sample Complexity of Pattern Classification with Neural Networks: The Size of the Weights is More Important than the Size of the Network. *IEEE Trans. Inf. Theory* **1998**, *44*, 525–536. [[CrossRef](#)]
41. Liu, X.; Gao, C.; Li, P. A comparative analysis of support vector machines and extreme learning machines. *Neural Netw.* **2012**, *33*, 58–66. [[CrossRef](#)] [[PubMed](#)]
42. Siqueira, H.; Boccato, L.; Attux, R.; Lyra, C. Echo state networks and extreme learning machines: A comparative study on seasonal streamflow series prediction. In Proceedings of the International Conference on Neural Information Processing, Doha, Qatar, 12–15 November 2012; pp. 491–500.
43. Lukosevicius, M.; Jaeger, H. Reservoir computing approaches to recurrent neural network training. *Comput. Sci. Rev.* **2009**, *3*, 127–149. [[CrossRef](#)]
44. Ozturk, M.C.; Xu, D.; Príncipe, J.C. Analysis and design of Echo State Networks. *Neural Comput.* **2007**, *19*, 111–138. [[CrossRef](#)]
45. Kulaf, A.C.P.; Von Zuben, F.J. Improved regularization in extreme learning machines. In Proceedings of the 11th Brazilian Congress on Computational Intelligence Porto de Galinhas, Pernambuco, Brazil, 8–11 September 2013; Volume 1, pp. 1–6.
46. Hashem, S. Optimal linear combinations of neural networks. *Neural Netw.* **1997**, *10*, 599–614. [[CrossRef](#)]
47. Siqueira, H.; Boccato, L.; Attux, R.; Lyra, C. Unorganized machines for seasonal streamflow series forecasting. *Int. J. Neural Syst.* **2014**, *24*, 1430009. [[CrossRef](#)]
48. Joe, H.; Kurowicka, D. *Dependence Modeling: Vine Copula Handbook*; World Scientific Publishing Co. Pte. Ltd.: Singapore, 2011.
49. Toly Chen, Y.C.W. Long-term load forecasting by a collaborative fuzzy-neural approach. *Int. J. Electr. Power Energy Syst.* **2012**, *43*, 454–464. [[CrossRef](#)]
50. CETESB-Environmental Sanitation Technology Company. Qualidade do ar no Estado de São Paulo, 2020. Available online: <https://cetesb.sp.gov.br/ar/publicacoes-relatorios> (accessed on 27 June 2021). (In Portuguese)
51. Datasus-Department of Informatics of the Unique Health System. SIHSUS Reduzida-Ministry of Health, Brazil. Available online: <http://www2.datasus.gov.br/DATASUS/index.php?area=0701&item=1&acao=11> (accessed on 1 July 2020).

52. IBGE-Brazilian Institute of Geography and Statistics (in Portuguese: Instituto Brasileiro de Geografia e Estatística. Censo 2010. 2021. Available online: <https://censo2010.ibge.gov.br/> (accessed on 27 July 2021).
53. Agrawal, S.B.; Agrawal, M. *Environmental Pollution and Plant Responses*; CRC Press: Boca Raton, FL, USA, 1999.
54. Tadano, Y.S.; Ugaya, C.M.L.; Franco, A.T. Methodology to assess air pollution impact on human health using the generalized linear model with Poisson Regression. In *Air Pollution-Monitoring, Modelling and Health*; InTech: São Paulo, Brazil, 2012.
55. Li, Y.; Ma, Z.; Zheng, C.; Shang, Y. Ambient temperature enhanced acute cardiovascular-respiratory mortality effects of PM 2.5 in Beijing, China. *Int. J. Biometeorol.* **2015**, *59*, 1761–1770. [[CrossRef](#)] [[PubMed](#)]
56. WHO-World Health Organization. *Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide-Global Update 2005-Summary of Risk Assessment, 2006*; WHO: Geneva, Switzerland, 2006.
57. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2021.
58. Haykin, S.S. *Neural Networks and Learning Machines*, 3rd ed.; Pearson Education: Upper Saddle River, NJ, USA, 2009.
59. Siqueira, H.; Boccato, L.; Luna, I.; Attux, R.; Lyra, C. Performance analysis of unorganized machines in streamflow forecasting of Brazilian plants. *Appl. Soft Comput.* **2018**, *68*, 494–506. [[CrossRef](#)]
60. Cuzick, J. A Wilcoxon-type test for trend. *Stat. Med.* **1985**, *4*, 87–90. [[CrossRef](#)] [[PubMed](#)]
61. Tadano, Y.S.; Potgieter-Vermaak, S.; Kachba, Y.R.; Chioli, D.M.; Casacio, L.; Santos-Silva, J.C.; Moreira, C.A.; Machado, V.; Alves, T.A.; Siqueira, H.; et al. Dynamic model to predict the association between air quality, COVID-19 cases, and level of lockdown. *Environ. Pollut.* **2021**, *268*, 115920. [[CrossRef](#)]
62. Khatri, K.L.; Tamil, L.S. Early detection of peak demand days of chronic respiratory diseases emergency department visits using artificial neural networks. *IEEE J. Biomed. Health Inform.* **2017**, *22*, 285–290. [[CrossRef](#)]
63. Moustris, K.P.; Douros, K.; Nastos, P.T.; Larissi, I.K.; Anthracopoulos, M.B.; Paliatsos, A.G.; Priftis, K.N. Seven-days-ahead forecasting of childhood asthma admissions using artificial neural networks in Athens, Greece. *Int. J. Environ. Health Res.* **2012**, *22*, 93–104. [[CrossRef](#)]
64. Shakerkhatibi, M.; Dianat, I.; Jafarabadi, M.A.; Azak, R.; Kousha, A. Air pollution and hospital admissions for cardiorespiratory diseases in Iran: Artificial neural network versus conditional logistic regression. *Int. J. Environ. Sci. Technol.* **2015**, *12*, 3433–3442. [[CrossRef](#)]
65. Cengiz, M.A.; Terzi, Y. Comparing models of the effect of air pollutants on hospital admissions and symptoms for chronic obstructive pulmonary disease. *Cent. Eur. J. Public Health* **2012**, *20*, 282. [[CrossRef](#)]
66. Zhou, R.; Wu, D.; Li, Y.; Wang, B. Relationship Between Air Pollutants and Outpatient Visits for Respiratory Diseases in Hangzhou. In Proceedings of the 2018 9th International Conference on Information Technology in Medicine and Education (ITME), Hangzhou, China, 19–21 October 2018; pp. 275–280.