IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 15, 2020, accepted December 29, 2020, date of publication January 11, 2021, date of current version January 26, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3050437

Neural-Based Ensembles for Particulate Matter Forecasting

PAULO S. G. DE MATTOS NETO[1], PAULO RENATO A. FIRMINO[2], HUGO SIQUEIRA[3], YARA DE SOUZA TADANO[4], THIAGO ANTONINI ALVES[5], JOÃO FAUSTO L. DE OLIVEIRA[6], MANOEL HENRIQUE DA NÓBREGA MARINHO[6], AND FRANCISCO MADEIRO[7]

[1]Centro de Informática, Universidade Federal de Pernambuco, Recife 50740-560, Brazil
[2]Center of Science and Technology, Federal University of Cariri, Juazeiro do Norte 63048-080, Brazil
[3]Department of Electrical Engineering, Federal University of Technology—Paraná, Ponta Grossa 84017-220, Brazil
[4]Department of Mechanical Engineering, Federal University of Technology—Paraná, Ponta Grossa 84017-220, Brazil
[5]Department of Mathematics, Federal University of Technology—Paraná, Ponta Grossa 84017-220, Brazil
[6]Polytechnic School of Pernambuco, University of Pernambuco, Recife 50720-001, Brazil
[7]UNICAP, ICAM-Tech International School, Catholic University of Pernambuco, Recife 50050-900, Brazil

Corresponding author: Paulo S. G. de Mattos Neto (psgmn@cin.ufpe.br)

This work was supported in part by Brazilian agencies—Brazilian National Council for Scientific and Technological Development, under Grant 405580/2018-5, Grant 308725/2015-8, Grant 315027/2018-5, and Grant 315298/2020-0; and in part by the Araucaria Foundation under Grant 51497.

ABSTRACT The air pollution caused by particulate matter (PM) has become a public health issue due to the risks to human life and the environment. The PM concentration in the air causes haze and affects the lungs and the heart, leading to reduced visibility, allergic reactions, pneumonia, asthma, cardiopulmonary diseases, lung cancer, and even death. In this context, the development of systems for monitoring, forecasting, and controlling emissions plays an important role. The literature about forecasting systems based on Artificial Neural Networks (ANNs) ensembles has been highlighted regarding statistical accuracy and efficiency. In this article, trainable and non-trainable combination methods are used for PM_{10} and $PM_{2.5}$ (particles with an aerodynamic diameter less than 10 and 2.5 micrometers, respectively) time series forecasting for eight different locations, in Finland and Brazil, for different periods. Trainable ensembles based on ANNs, linear regression, and Copulas are compared with non-trainable combinations (mean and median), single ANNs, and linear statistical approaches. Different models are considered so far, including Autoregressive model (AR), Autoregressive and Moving Average Model (ARMA), Infinite Impulse Response Filters (IIR), Multilayer Perceptron (MLP), Radial Basis Function Networks (RBF), Extreme Learning Machines (ELM), Echo State Networks (ESN), and Adaptive Network Fuzzy Inference System (ANFIS). The use of ANNs ensembles, mainly combined with MLP, leads to a better one step ahead forecasting performance. The use of robust air pollution forecasting tools is prime to assist governments in managing air pollution issues like hospital collapse during adverse air quality situations. In this sense, our study is indirectly related to the following United Nations sustainable development goals: SDG 3 - good health and well-being and SDG 11 - sustainable cities and communities.

INDEX TERMS Forecasting, particulate matter, artificial neural networks, ensemble.

I. INTRODUCTION

Air pollution is one of the worst toxic issues worldwide [1]–[3]. The World Health Organization (WHO) [4] reported that 0.8 million deaths and 7.9 million disability-adjusted life years from respiratory problems, lung diseases, and cancer were attributed to urban air pollution. More recently, WHO [5] reported that 90% of the urban population is exposed to high air pollution levels. Many works, including several epidemiological studies, have reported the relationship between particulate matter concentration and cardiorespiratory diseases and even death [6]–[12].

Particles with an aerodynamic diameter less than 10 μm (PM_{10}), and mainly those less than 2.5 μm ($PM_{2.5}$), pose severe damages to the environment and human health.

Environmental damages [2], [3] may include pollution and acidity of lakes and rivers, imbalance in coastal water and large river basins, depletion of soil, acid rain, and damaging to forests, farm crops, and ecosystems.

The associate editor coordinating the review of this manuscript and approving it for publication was Haiyong Zheng.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

Human health damages commonly vary according to the level of concentration of PM, the period of exposure, size of the particulates, and the atmospheric chemical profile at a location [3], [9]. In general, short-term exposure causes less severe damages, like allergic reactions and irritations in the upper respiratory tract. In some cases, more serious problems [13] may occur, like emphysema, pneumonia, and asthma. In the same way, long-term exposure may severely affect the respiratory system and other parts of the human body [14], [15], such as the brain, liver, kidneys, and cardiovascular system; causing heart disease, cancer, chronic lower respiratory diseases, cerebrovascular disease, and even death.

Due to the risks related to the PM_{10} and $PM_{2.5}$ concentration in air, it is crucial that government agencies of air quality alert in advance the population about the onset, severity, and duration of high concentration episodes. These institutional actions mainly aim: (i) in the short-term to draw attention to harmful effects of high PM concentration [2], [3] and (ii) in the long-term, to encourage the population and industry to reduce emissions of PM [2], [3]. In this context, it has been paramount the PM concentration monitoring and forecasting.

Several forecasting systems [16]–[22] have been developed to reach good estimates of the PM concentration. In the literature, those based on Artificial Neural Networks (ANNs) have been highlighted due to their performance and generalization capability.

However, studies like Neumann's [23], comment that adopting a single model can lead to statistical bias and underestimating the real uncertainty underlying the time series. In this way, authors [24]–[30] have been challenged by combining diverse models to present aggregate estimates.

In this context, one of the problems has been to combine single predictors of the time series to enhance the forecasting performance. Statistically, these ensembles are generally superior in comparison with individual models in terms of both accuracy and efficiency [24], [31]–[34]. Also, among ANN-based approaches, there are ensembles that combine models to obtain a more statistically robust system, outperforming the single forecasting models [12], [29], [35], [36].

The combination methods employed in the ensembles can be divided into two classes: non-trainable [36] and trainable [29]. The non-trainable combination operators commonly used are descriptive statistics, such as mean, median, mode, maximum, and minimum [36]. In turn, the trainable combination methods require a phase of prior estimation of parameters that aims to find the best function of aggregation of the forecasts [29].

The method or operator used in the combination is thus paramount for developing attractive models. Currently, several works in the PM forecasting literature use this kind of approach. For instance, Siwek and Osowski [37] improved different types of PM_{10} models by using a wavelet transformation jointly with an ensemble employing two approaches, separately: Support Vector Regression and Multilayer Perceptron; Souza *et al.* [38] proposed an ensemble of ANNs based on bagging to predict the daily concentrations of PM_{10} in the city of Piracicaba, Brazil, while Debry and Mallet [39] proposed a method named discounted ridge regression (DRR) to combine machine learning algorithms for prediction of PM_{10} concentration in France.

Generally, the forecasting models that can or cannot be associated to an ensemble generally employ two data-driven approaches for their training: the use of only previous (historical) data of the PM concentration [16]–[18], or the use of the historical data of PM jointly with related features, such as temperature, relative humidity, direction, and speed of the wind [37]. Several works in the literature have considered the second approach [19]–[22], [37], [39].

There is no consensus on which model shows better performance to each problem. Then, to develop and test different approaches is crucial to keep improving air pollution forecasting. In such a context, the present work investigates the performance of 16 forecasting models for PM concentration. As single models we applied the Autoregressive model (AR) [40], Autoregressive and Moving Average model (ARMA) [40], Infinite Impulse Response Filters [41], Multilayer Perceptron (MLP) [20], Radial Basis Function Networks (RBF) [42], Extreme Learning Machines (ELM) [43], Echo State Networks (ESN) [44], and Adaptive Network Fuzzy Inference System (ANFIS) [45]; as non-trainable ensembles, it is considered the mean and median [36]; as trainable ensembles we addressed a linear regression with (LR-FS) and without features selection (LR) [46], ELM with and without the coefficient of regularization (CR), MLP [37], and normal Copula-based models [25], [29]. These combination methods have emerged among the most promising from the time series forecasting literature [12], [29], [35], [37], [38].

Thus, the approaches are evaluated via historical PM series in terms of six performance metrics widely used in the literature: Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), Average Relative Variance (ARV), Index of Agreement (IA), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

The target was PM_{10} and $PM_{2.5}$ concentrations time series. As a geographical location, socioeconomic factors, and urban development strategy highly impact the air quality in the cities [47], we studied eight different scenarios for different geographical locations from Brazil and Finland. According to the authors' knowledge, comparing this set of models in PM forecasting tasks is unprecedented.

The rest of the work is organized as follows: Section II presents the background regarding the single and combination methods adopted for modeling and forecasting PM; Section III presents computational results, while Section IV brings relevant discussions; Section V shows concluding remarks.

II. BACKGROUND

A. LINEAR FORECASTING MODELS

Linear forecasting models are traditional statistical tools to perform time series modeling and forecasting [40]. Besides

developing nonlinear methodologies, the literature shows that such an approach is widely used [11], [48]–[50]. In this sense, we present three of the most used linear models to deal with the aforementioned task. In this way, let x_t be the value of the time series under consideration at instant t.

1) AUTOREGRESSIVE MODELS

Autoregressive (AR) models are popular methods for stationary time series prediction [40] due to its simplicity in optimizing its parameters and implementing, allied to good results in the literature. The model weights previous values of the time series to predict its future values. Equation 1 summarizes the formalism:

$$\hat{x}_t = \phi_1 x_{t-P} + \phi_2 x_{t-P-1} + \cdots + \phi_p x_{t-P-p+1} + a_t \quad (1)$$

in which $x_{t-P-i+1}$ $(i = 1, 2, \ldots, p)$ are the lags of the observed series, ϕ_i represents the free parameters, and a_t are random shocks (or the random component) [48]. Thus, p is the order of the model.

Equation 1 allows one to promote P steps ahead forecasts [51]. The model presents a unimodal cost function in the MSE sense. Therefore, the global minimum is defined by a closed-form solution, named Yule-Walker equations [41].

2) AUTOREGRESSIVE AND MOVING AVERAGE MODELS

The Autoregressive and Moving Average Models (ARMA) is developed as a hybrid model between the AR and the Moving Average (MA) models. While the AR considers the lags of the series, the ARMA also creates the output response addressing the previous residuals presented by the model, a_{t-P-j}, which are weighted by θ_j coefficients, as in Equation 2:

$$\begin{aligned}\hat{x}_t = {} & \phi_1 x_{t-P} + \cdots + \phi_p x_{t-P-p+1} \\ & - \theta_1 a_{t-P} - \cdots - \theta_q a_{t-P-q+1} + a_t \end{aligned} \quad (2)$$

in which ϕ_i, $i = 1, 2, \ldots, p$ and θ_j, $j = 1, 2, \ldots, q$, are free coefficients [40].

The standard application of ARMA considers the random shock a_{t-P-j} as equivalent to the residuals of the previous samples [40], [48]. The feedback of previous temporal information is the reason classify the model as a recursive linear approach. However, the estimation of the ARMA parameters has no closed-form solution, being necessary the use of probabilistic optimization methodologies to adjust the model. Therefore, it may be unfeasible to carry out an exhaustive search for the ARMA best coefficients.

In this sense, we use a bio-inspired metaheuristic to adjust the ARMA model: the well known Particle Swarm Optimization Algorithm (PSO) [52]–[55].

3) INFINITE IMPULSE RESPONSE FILTERS

Recursive linear models can be described from a different perspective of the ARMA models. Instead of feeding back previous residual values, one can reinsert previous values of the model output [41], [56]. In this case, the model is known as the Infinite Impulse Response Filter (IIR), which is depicted in Equation 3 [41]

$$\begin{aligned}\hat{x}_t = {} & c_1 x_{t-P} + \cdots + c_p x_{t-P-p+1} \\ & - b_1 \hat{x}_{t-P} - \cdots - b_q \hat{x}_{t-P-q+1} \end{aligned} \quad (3)$$

in which c_i, $i = 1, 2, \ldots, p$, are free parameters that weigh the feedforward inputs and b_j, $j = 1, 2, \ldots, q$, are the weights of the feedback inputs.

As in the ARMA case, the IIR Filters cannot be adjusted by closed-form solutions [41]. Again, we use the PSO to estimate the parameters of the model.

B. ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (ANNs) are nonlinear methodologies, inspired by the functioning of the superior organisms' neural system. Composed by nonlinear processing structures named artificial neurons, such models are universal approximators, with high mapping capability [42]. The adjustment of an ANN is known as the training process [57].

This class of methods is vastly applied to time series forecasting [58]–[60], and nonlinear mapping problems [9], [61]. In this work, we address four ANN frameworks: MLP, RBF, ELM, and ESN, sumarized as follows.

1) MULTILAYER PERCEPTRON - MLP

Feedforward neural networks (FNN) are those in which the information signal flows in one direction, from the input layer to the output. These architectures present universal approximation capability, which means that they can approximate any continuous, nonlinear, limited, and differentiable function [42].

A popular FNN is the Multilayer Perceptron (MLP). The most widely known method for adjusting the weights of the MLP is the backpropagation algorithm [42]. Here, we use an MLP with three layers, with hyperbolic tangent and linear function as activation functions for the hidden and output layers, respectively. However, in this work the training process is performed using the Modified Scaled Conjugated Gradient [62].

2) RADIAL BASIS FUNCTION NETWORKS - RBF

The Radial Basis Function Networks (RBF) is a feedforward ANN framework that presents two layers, hidden and output layers. The hidden layer performs an input-output mapping, using radial basis functions as kernel (activation) ones [63]. The most used function is the Gaussian. In this case, the artificial neurons have two free parameters, a center and a dispersion. The output layer provides a combination of the hidden layer outputs, often using a linear approach [42].

The training step of an RBF is performed according to two stages. The first is the determination of the hidden layer weights, in which one must determine their centers and dispersions. This task is performed using non-supervised clustering methods. In this work, we address the K-Medoids methodology [64].

The second step is to tune the weights of the output layer. In this work, we address the Moore-Penrose pseudoinverse operator, a linear procedure, which ensures the minimum mean squared error (MSE) in the output [60].

3) EXTREME LEARNING MACHINES - ELM

Extreme Learning Machines (ELM) are single layer feed-forward neural networks, introduced by Huang *et al.* [43] in 2004. The arrangement of the neurons is quite similar to the architecture of the MLP. However, ELM presents a remarkable dissemblance in the training process. It has been proven that the weights in the hidden neurons, which are randomly generated, can stand untuned. For this, the activation functions of the intermediate neurons must be continuously differentiable [63]. The authors proved that the insertion of new neurons in this layer leads to decreased output error. The ELM are universal approximators [65].

The training process of an ELM is simple, being summarized in finding the best set of weights of the neurons in the output layer. This task can be performed by solving a linear regression problem. Huang *et al.* [65] suggest using the Moore-Penrose generalized inverse operation to solve the task since this technique simultaneously minimizes the norm of the output weight vector and the MSE between the output of the network and the desired signal [66].

To increase the generalization capability of an ELM, we can also address the coefficient of regularization (CR) [35], [57].

4) ECHO STATE NETWORKS - ESN

Jaeger [44] proposed the Echo State Networks (ESN), which present similarities with the ELM in terms of the simplicity in the training process.

The most important structural difference between ELM and ESN is the presence of recurrent connections within the latter's intermediate layer, called dynamic reservoir. The neurons' activation in the reservoir (the output of this layer) is influenced by the current input and the previous state.

Under specific conditions, the reservoir output is a nonlinear transformation, directly influenced by the recent history of the input signal (hence the term echo). It allows this layer to be set in advance and kept unchanged during the training. Therefore, only the output layer must be adjusted through a solution of a least-square problem. He called these conditions as *echo state PROPERTY* [67].

Jaeger suggests a way to generate the reservoir that respects the echo state property. In his proposal, he created a sparse matrix. Also, as in the ELM case, the training process is performed applying the Moore-Penrose inverse operation [51].

C. NEURO-FUZZY MODEL

Adaptive Network Inference Fuzzy System (ANFIS) [68] is a hybrid neuro-fuzzy-based model that combines artificial neural networks and fuzzy logic. ANFIS carries out a cross-validation by data set checking, leading to minimization of overfitting occurrence. It was previously applied to particulate matter prediction [69], [70].

It is widely known that the number of ANFIS parameters increases exponentially with the number of input variables (which corresponds in the present work to the number of past PM sample values used for prediction), being an unfavorable point in comparison to traditional approaches of ANN. However, the versatility of ANFIS suggests that it may succeed in cases that neural nets have failed.

The ANFIS model uses the following parameters [71]:

- *TMF*: type of membership functions;
- *NMF*: number of membership functions;
- *NI*: number of inputs;
- *NFR*: number of fuzzy rules.

One can observe that

$$NFR = NMF^{NI}. \tag{4}$$

The following membership functions were considered in the present work:

- *trimf*: Triangular membership function;
- *gauss2mf*: Two-sided Gaussian membership function;
- *dsigmf*: Membership function given by the difference between two sigmoid membership functions;
- *pimf*: Pi-shaped curve membership function;
- *gaussmf*: Gaussian curve membership function;
- *gbellmf*: Generalized Bell curve membership function;
- *trapmf*: Trapezoidal membership function;
- *psigmf*: Product of two sigmoid membership functions.

D. COMBINATION MODELS

1) ENSEMBLES

The ensemble methodology combines the output of single forecasting models to improve the final response of the system [37]. The underlying reasoning is that distinct formalisms can deal with the diverse characteristics of the time series, even in the light of the same set of inputs. In this sense, one method may present better responses for some data range, while other models may work better in another band. Therefore, a combination approach can be used to generate the final output [36].

We highlight the necessity of accurate predictions of each single model, as well as they have to present diversity [72]. Efficiency also plays an important role. In the present work, ($n = 8$) single models are considered: AR, ARMA, IIR Filter, MLP, RBF, ELM, ESN, and ANFIS. As discussed, such models are of a different nature, among linear models, neural networks, and neuro-fuzzy approaches, to reach diversity in the final response [38], [39].

We addressed as combination formalisms:

- non-trainable approaches: mean and median;
- trainable approaches: linear regression (LR), linear regression with feature selection (LR-FS), MLP, ELM, ELM (CR), and normal copula-based models (NC).

The outputs combination through an LR model [46], [73] is based on a linear combination of the predictions obtained

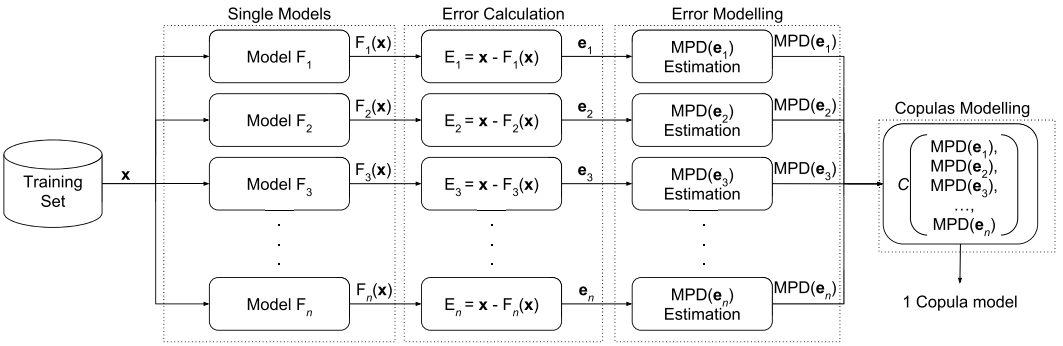

FIGURE 1. Estimation steps for the copula model.

from the single models as shown in Equation 5,

$$x_{t+1}^{LR} = \sum_{j=1}^{n} w_j^{LR} F_j(\mathbf{x}_t), \tag{5}$$

in which x_{t+1}^{LR} is the combined output, n is the number of single models, w_j^{LR} is the coefficient that weight the output of the j_{th} single model forecast for x_{t+1}, $F_j(\cdot)$, regarding the observed series until t, $\mathbf{x}_t$.

The weights w_j are calculated by means of a least squares approach, using the predictions of each model and the target value. The weights reflect the contribution of each model in the final result.

Feature selection is also employed using the LR approach, where a wrapper [74], [75] is used to find the best set of models to be considered in a given data set. Thus, a search is conducted over $2^8 - 1$ possible combinations.

2) NORMAL COPULA-BASED COMBINATION

A Copula is a function that combines two or more univariate marginal probability distributions (MPDs) to build a joint probability distribution (JPD), incorporating the dependence of these univariate distributions [76]. In this context, a marginal cumulative distribution function (CDF) can be seen as a MPD and a joint CDF can be understood as a JPD [77].

In the time series context, the use of copulas for economics and finance has been paramount for modeling the dependence of variables through time [78], [79]. Works have studied the introduction of copulas formalism in order to achieve maximum likelihood combination models according to an adequate JPD [29], [80]–[82].

In general terms, a copula function, $\mathcal{C}(\cdot)$, is a JPD whose marginal distributions are in the range [0,1]. Let $(v_1, \cdots, v_j, \cdots, v_n)$ be an instance of n MPDs such that $v_j \in [0, 1]$. Then, the copula probability density function (PDF) is given by Equation 6

$$c\,(v_1, \cdots, v_n) = \frac{\partial^n}{\partial v_1 \cdots \partial v_n} \mathcal{C}\,(v_1, \cdots, v_n)\,, \tag{6}$$

in which $\mathcal{C}(v_1, \cdots, v_j, \cdots, v_n)$ is the respective JPD [76].

Figure 1 summarizes the flow for a copulas-based combination of n single models, $(F_1(\mathbf{x}), \cdots, F_n(\mathbf{x}))$, according to a training series $\mathbf{x}$. In a divide-and-conquer way, the residuals of each single forecasting model $(\mathbf{e_1}, \cdots, \mathbf{e_n})$ are modeled and encapsulated in MPDs. In the general formulation, v_j can be an instance of the j^{th} MPD in the light of a fixed residual e_j. Then, such MPDs are copulated.

This article focuses on the Normal or Gaussian copula. This copula belongs to the family of Elliptic copulas. In this case, the dependence between pairs of variables is given by an $n \times n$ covariance matrix, Σ, from which a correlation matrix, ρ, can be obtained. For instance, when $\rho_{i,j}$ is the Pearson (linear) correlation between the variables of indexes i and j, $\rho_{i,j} \in [-1, 1]$, it follows that $\rho_{i,j} = 0$ indicates no (linear) correlation between the variables, $\rho_{i,j} = -1$ corresponds to perfect linear negative correlation, and $\rho_{i,j} = 1$ reflects perfect linear positive correlation.

The Normal copula is given by Equation 7 [83]:

$$\mathcal{C}(v_1, \cdots, v_n) = \Phi(\varphi^{-1}(v_1), \cdots, \varphi^{-1}(v_n)|\rho), \tag{7}$$

in which $\varphi^{-1}(\cdot)$ is the inverse MPD of a standard normal distribution and $\Phi(\cdot)$ is the JPD of a multivariate normal distribution with zero mean vector and correlation matrix equal to ρ.

The corresponding PDF of the Normal copula is given in Equation 8 [83]

$$\begin{aligned} c(v_1, \cdots, v_n) &= c(\varphi^{-1}(v_1), \ldots, \varphi^{-1}(v_n)|\rho) \\ &= \frac{1}{\sqrt{|\rho|}} exp\left(-\frac{1}{2}(\varphi^{-1}(v_1), \right. \\ &\qquad \left. \cdots, \varphi^{-1}(v_n))(\rho^{-1} - \mathbb{I}) \begin{pmatrix} \varphi^{-1}(v_1) \\ \vdots \\ \varphi^{-1}(v_n) \end{pmatrix}\right) \end{aligned} \tag{8}$$

in which $\mathbb{I}$ is the identity matrix and Σ^{-1} is the inverse correlation matrix.

For the case where v_j comes from a normal MPD, the resulting copula model is the multivariate Normal distribution. It is the most widely used multivariate model in

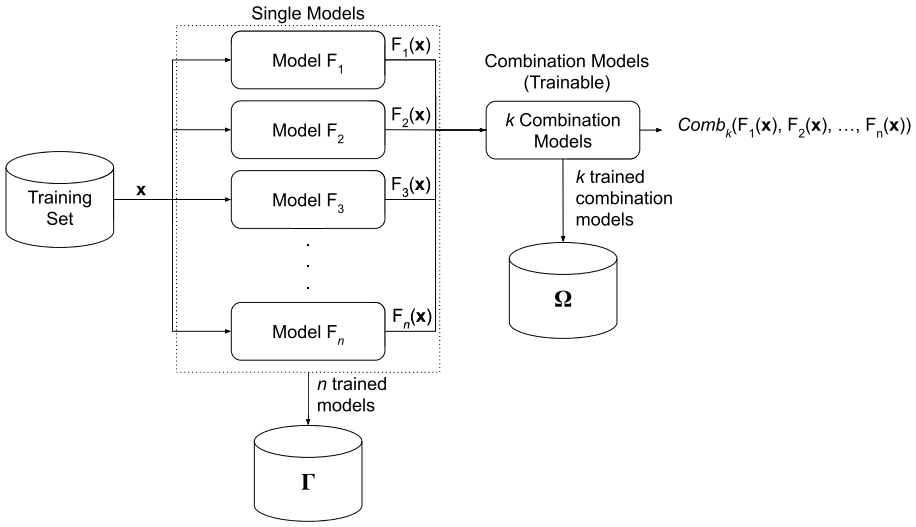

FIGURE 2. Training step for the *k* trainable combination models (ensemble).

statistics. The multivariate Normal distribution has been useful for principal component analysis, classical regression, and ensembles (as used in [25]), for instance. Further, in this case, the maximum likelihood combination estimate from copulas-based approach is similar to the minimal variance (MV) method [25], [84]. Thus, this combination approach is named Normal Copula-based (NC) hereafter.

The NC promotes linear combination forecasts (LCF) $\mathbf{x}^{\mathbf{NC}}$ and can be generically presented as in Equation 9:

$$x_{t+1}^{NC} = \sum_{j=1}^{n} \omega_j F_j(\mathbf{x}_t), \tag{9}$$

in which n is the number of single models and ω_j is the weight attributed to the j^{th} single model forecast for x_{t+1}, $F_j(\cdot)$, in the light of the previous series $\mathbf{x}_t$.

In NC approach, ω_j is a function of both the efficiency and linear correlation of the single models, as presented in Equation 10 [25]:

$$\omega_j = \frac{\sum_{l=1}^{n} h_{lj}}{\sum_{l=1}^{n}\sum_{j=1}^{n} h_{lj}}, \tag{10}$$

in which h_{lj} is the j^{th} element of the l^{th} row of the inverse covariance matrix Σ^{-1} of the residuals of the single models.

Thus, the greater the dependence between two models and the greater the absolute magnitude of their residuals, the lesser their weights in the combination.

For instance, if $n = 2$, we have Expression 11:

$$\Sigma^{-1} = \frac{1}{1-\rho_{1,2}^2} \cdot \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho_{1,2}}{\sigma_1\sigma_2} \\ -\frac{\rho_{1,2}}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2}, \end{pmatrix}, \tag{11}$$

in which $\rho_{1,2} = \sigma_{12}/\sigma_1\sigma_2$ is the Pearson correlation coefficient between the errors of the models $F_1(\cdot)$ and $F_2(\cdot)$, σ_i is the standard deviation of the error of $F_i(\cdot)$, and $\sigma_{i,j}$ is the covariance between the errors of $F_i(\cdot)$ and $F_j(\cdot)$.

III. CASE STUDIES

A methodology may bring gains to a specific time series but not to others. Then, we studied the databases from two countries with distinct climate and emission patterns (Brazil and Finland) to analyze the performance of each method under study.

Figures 2 and 3 illustrate the general idea of the work for studying the performance of single and combination models for PM time series forecasting. From the observed series, i.e., the training data set ($\mathbf{x}$), the use of n single models ($F_1(\mathbf{x}), \cdots, F_n(\mathbf{x})$) is suggested. Such single models are then combined according to the trainable approaches (ensembles and Copulas), taking into account $\mathbf{x}$ (training data). Figure 2 shows a scheme of the training process for single and combination models.

After this, the single and combination models are used for forecasting data outside the training set, as shown in Figure 3. Therefore, for a given vector of time lags x_q, which belongs to the PM series test set, the one step ahead forecasts are performed.

A. PM TIME SERIES

The database of PM_{10} and $PM_{2.5}$ concentrations addressed in this work consists of univariate time series composed of daily mean records to four different cities, with distinct characteristics:

- Helsinki, Finland - Kallio and Vallila stations - PM_{10} and $PM_{2.5}$;
- São Paulo city, São Paulo state, Brazil - Tietê station - PM_{10} and $PM_{2.5}$;
- Campinas city, São Paulo state, Brazil - PM_{10};
- Ipojuca city, Pernambuco state, Brazil - PM_{10}.

Helsinki is the most populous city and the capital of Finland, with 655,276 inhabitants spread over a 1,268,296 km^2

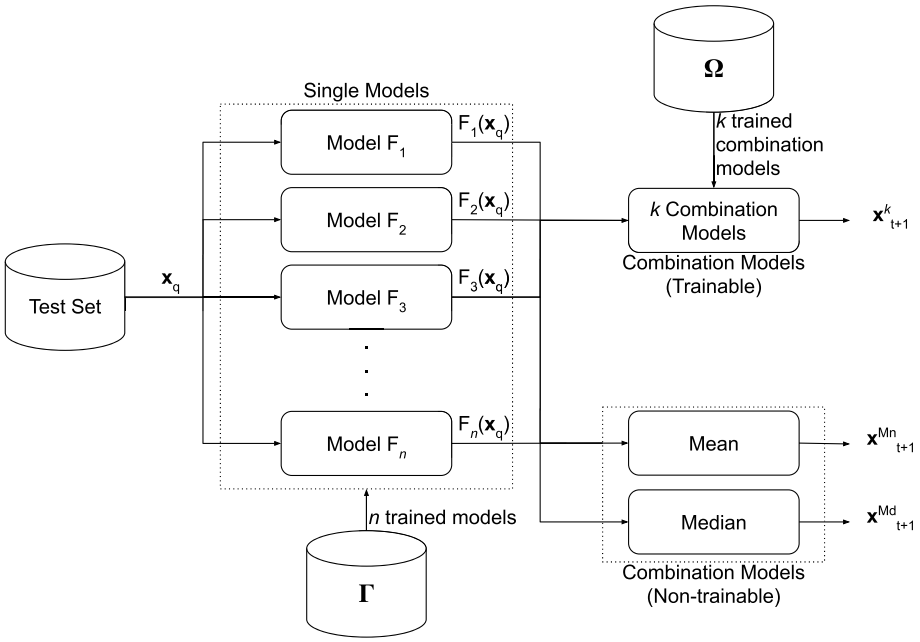

FIGURE 3. Test step for the proposed ensemble. The outputs x^{k}_{t+1}, x^{Mn}_{t+1}, and x^{Md}_{t+1} are the forecasts of the combinations for a given pattern test x_q.

of area [85]. It has temperatures ranging from −8°C to 21°C, and hardly below −19°C and above 26°C [86]. Helsinki time series have already been addressed in the literature about air pollution forecasting [17]–[20], [87].

In contrast to Helsinki, Ipojuca city (Pernambuco state, Brazil) weather is characteristic of hot and windy climate with temperatures ranging from 22°C to 32°C [86]. Ipojuca is a coastal city located in the Brazilian Northeast, Pernambuco state capital, and has a demographic density of 152.98 inhabitants per km^2 (total population of 96,204 inhabitants) [88].

The two cities of São Paulo state (São Paulo and Campinas) also have a distinct population and business characteristics, but similar weather (hot and rainy long summers and short winters with temperatures raging from 13 °C to 29°C [86]). They are located in Southeast Brazil. São Paulo is the capital of São Paulo state and the most populous city in Brazil (7398,26 inhabitants per km^2 - total population of 12,252,023 and 969.32 km^2 of urban area) [88].

Campinas city demographic density is five times lower than São Paulo city (1359.60 inhabitants per km^2), with an urban area of 238.2 km^2 [88].

Beyond the diverse climate, demographic, and business characteristics of the studied locations, each monitoring location has a distinct emission source pattern. As shown in Figure 4, Vallila station is in a high traffic, city downtown area. In contrast, Kallio station is an urban background in Helsinki has quite peculiar demographic, business, and climatic characteristics, compared to Brazilian areas under study.

Figure 5 shows São Paulo, Campinas, and Ipojuca cities stations' location. São Paulo and Campinas cities are dominated by vehicular sources. The difference is that São Paulo city's monitoring region is near a ring road, with main influence of heavy-duty vehicles, while Campinas station is in the city downtown, predominantly affected by light-duty emissions. Ipojuca city is quite distinct, as it is a coastal city. The main difference from the other time series is that the monitoring station is located near a petrol refinery, being characterized by industrial emissions. A statistical description of all series is shown in the Tables 1 and 2.

TABLE 1. Number of samples, data range, and considered pollutant to each studied station.

Station	N.Samples	Time Range
Kallio–PM_{10}	1090	Jan 1st 2001 to Dec 31st 2003
Kallio–$PM_{2.5}$	1095	Jan 1st 2001 to Dec 31st 2003
Vallila–PM_{10}	1092	Jan 1st 2001 to Dec 31st 2003
Vallila–$PM_{2.5}$	961	Jan 1st 2001 to Dec 31st 2003
São Paulo–PM_{10}	1095	Jan 1st 2017 to Dec 31st 2019
São Paulo–$PM_{2.5}$	1095	Jan 1st 2017 to Dec 31st 2019
Campinas–PM_{10}	731	Jan 1st 2007 to Dec 31st 2008
Ipojuca–PM_{10}	632	Jul 17th 2015 to Apr 9th 2017

TABLE 2. Mean, standard deviation, maximum, and minimum values for each studied station.

Pollutant	Station	Mean	S. Deviation	Max	Min
PM_{10} [μg/m^3]	Kallio	16.7	9.9	80.0	2.9
	Vallila	20.3	12.7	138.0	3.1
	São Paulo	32.1	17.3	102.4	5.5
	Campinas	38.0	15.2	128.7	12.2
	Ipojuca	35.8	11.1	78.7	2.8
$PM_{2.5}$ [μg/m^3]	Kallio	8.8	5.5	56.8	1.8
	Vallila	9.9	6.1	70.5	1.9
	São Paulo	19.7	11.3	64.1	3.4

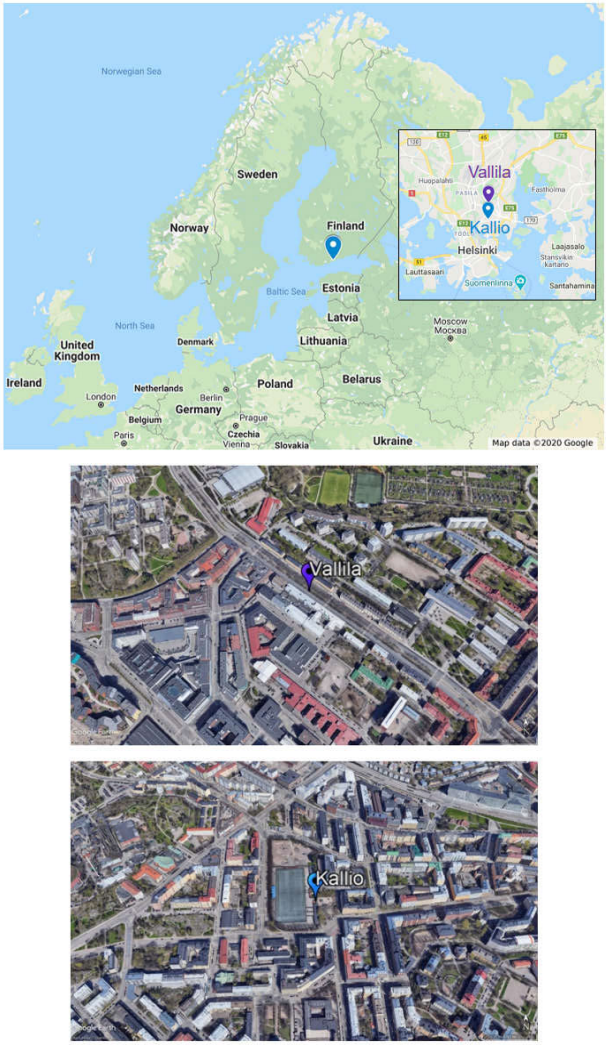

FIGURE 4. Location of Finland Stations at Helsinki: Kallio (blue spot) and Vallila (purple spot). The satellite map is from Google Maps (Map data 2020 Google; https://www.google.com/maps/place/Finland/); the satellite is from Google Earth Pro (Map data 2020 Google; https://www.google.com/maps/@60.1834096,24.8975655,12.83z).

FIGURE 5. Location of Brazilian Stations: Ipojuca (green spot), Campinas (yellow spot), and São Paulo (red spot). The satellite map is from Google Maps (Map data 2020 Google, INEGI; https://www.google.com/maps/place/Brazil/); the satellite is from Google Earth Pro (Map data 2020 Google; https://www.google.com/maps/@-23.2636702,-47.1095854,9.5z and https://www.google.com/maps/@-8.0624551, -34.9114682,11.92z).

B. EXPERIMENTAL SETUP

Initially, the PM concentration series were normalized using the z-score [11], [40] and then normalized to lie within the interval $[-1, 1]$ [42]. The samples were divided into three sets, according to Proben [89]: 50% for training, 25% for validation, and 25% for test. The test sets comprise the last samples of each series: 272 samples to Kallio (PM_{10} and $PM_{2.5}$), Vallila (PM_{10}), São Paulo (PM_{10} and $PM_{2.5}$); 240 to Vallila ($PM_{2.5}$); 182 samples to Campinas (PM_{10}); and 158 samples to Ipojuca (PM_{10}).

Thirty simulations with each artificial neural network were performed, and the best configuration was selected according to the lowest MSE value in the validation set. The selection of the best set of inputs (lags) is defined in every single model through the wrapper method [74], [75]. As the methodology is model-dependent, the forecasting models can select different lags.

The AR model was adjusted using the Yule-Walker equations, while the ARMA and IIR Filter were tuned by the PSO algorithm. The target during the training was to minimize the cost function based on the MSE [12].

The MLP is trained via the Modified Scale Conjugated Gradient [62] algorithm using the following stopping conditions: (i) the maximum number of iterations equal 300; (ii) use of the hold-out cross-validation; (iii) progress training 10^6. The number of hidden neurons for a single MLP model is defined using a grid search in the range $[3, 250]$. For the MLP used in the combination, the number of input nodes is set up eight, one for each single ANN model. The hyperbolic tangent is addressed as the activation function of the hidden neurons. The RBF was configurated following the same premisses of the MLP.

In addition to using grid partition, ANFIS uses a hybrid optimization method: the combination of least-squares estimator and backpropagation as gradient descent. The product operator was used as the connective, and the weighted mean

for defuzzification. A zero tolerance for the error is adopted in the stopping criterion, and the number of iterations adopted is 250. The number of inputs (NI) is selected in the range $[1, 6]$, and the number of fuzzy rules (*NFR*) is selected into the interval $[2, 50]$. The type of member functions (*TMF*) is selected among eight candidates: *trimf, gauss2mf, dsigmf, pimf, gaussmf, gbellmf, trapmf* and *psigmf*.

The weights of ELM and ESN are adjusted by the Moore-Penrose inverse operation [42], [57]. To determine the number of neurons in the hidden layer of the neural models, a grid search is carried out in the interval $[3, 250]$. Both also use the hyperbolic tangent as activation function of the hidden neurons.

Six metrics are taken into account for evaluating the single and combination PM models [90], [91]: MSE, Mean Absolute Percentage Error (MAPE), Average Relative Variance (ARV), Index of Agreement (IA), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE), given by Equations 12 to 17:

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(x_t - \hat{x}_t)^2, \tag{12}$$

$$MAPE = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{x_t - \hat{x}_t}{x_t}\right|, \tag{13}$$

$$ARV = \frac{\sum_{t=1}^{N}(x_t - \hat{x}_t)^2}{\sum_{t=1}^{N}(\hat{x}_t - \overline{x})^2}, \tag{14}$$

$$IA = 1 - \frac{\sum_{t=1}^{N}(x_t - \hat{x}_t)^2}{\sum_{t=1}^{N}(|\hat{x}_t - \overline{x}| + |x_t - \overline{x}|)^2}, \tag{15}$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|x_t - \hat{x}_t|, \tag{16}$$

$$RMSE = \sqrt{MSE}, \tag{17}$$

in which N is the number of available samples, x_t is the actual value of the series at time index t, $\hat{x}_t$ is the model forecast for x_t, and $\overline{x}$ the mean of the series.

In the case of the MSE, MAPE, MAE, ARV, and RMSE measures, the lower the value of those measures, the better the performance of the model. Mainly, ARV is a metric used to compare the methodologies with the simple mean of the series. If the ARV value is 1, the prediction of the model is as good as using the mean as the prediction of the series; otherwise, if the value is less (greater) than 1, the prediction of the model is better (worse) than using the mean as the prediction. In turn, the higher the IA, the better the model. Via IA, one can evaluate the quality of the model concerning both the accuracy of the simple mean estimate and the dispersion of the series [19], [20], [87].

TABLE 3. Evaluation metrics for forecasting PM_{10} series (Kallio Station).

	Model	MSE	MAPE	ARV	IA	MAE	RMSE
Single	AR	38.64	37.14	**1.00**	0.78	4.49	6.22
	ARMA	36.34	37.44	1.45	0.80	4.36	6.03
	IIR Filter	35.56	**36.64**	1.33	0.80	**4.30**	5.96
	RBF	36.62	38.88	1.36	0.79	4.40	6.05
	MLP	**34.44**	37.94	1.23	**0.80**	4.40	**5.87**
	ELM	34.90	37.80	1.33	0.80	4.34	5.91
	ESN	35.30	37.65	1.17	0.80	4.38	5.94
	ANFIS	37.22	38.48	1.52	0.79	4.41	6.10
Combination	Mean	35.30	37.30	1.29	0.80	4.32	5.94
	Median	35.25	37.53	1.31	**0.80**	4.33	5.94
	LR	35.87	37.85	1.38	0.80	4.43	5.99
	LR (FS)	35.57	37.39	1.26	0.79	4.32	5.96
	ELM	33.74	36.64	1.09	0.79	4.31	5.81
	ELM (CR)	33.84	37.10	**1.08**	0.79	4.33	5.82
	MLP	**32.37**	**29.43**	1.09	0.79	**4.08**	**5.69**
	Copulas	34.99	37.80	1.33	0.80	4.39	5.92

TABLE 4. Ranking of the single and combination models by metric for PM_{10} series of the Kallio Station.

	Model	MSE	MAPE	ARV	IA	MAE	RMSE	Mean	Rank
Single	AR	16	5	1	16	16	16	11.67	13
	ARMA	13	8	15	8	9	13	11	12
	IIR Filter	10	3	12	4	2	10	6.83	4
	RBF	14	16	13	13	13	14	13.83	15
	MLP	4	14	6	1	12	4	6.83	4
	ELM	5	12	10	2	8	5	7	7
	ESN	9	10	5	7	10	9	8.33	10
	ANFIS	15	15	16	14	14	15	14.83	16
Combination	Mean	8	6	8	6	5	8	6.83	4
	Median	7	9	9	3	7	7	7	7
	LR	12	13	14	9	15	12	12.5	14
	LR (FS)	11	7	7	10	4	11	8.33	10
	ELM	2	2	4	11	3	2	4	2
	ELM (CR)	3	4	2	12	6	3	5	3
	MLP	1	1	3	15	1	1	3.67	1
	Copulas	6	11	11	4	11	6	8.17	9

We highlight that, after the forecasting procedure, we applied the Friedman test to evaluate if the performances were statistically different [92]. Considering 5% of significance, the highest p-value found was 1.3914e-73, which allows admitting that a change in the predictor led to different results.

C. COMPUTATIONAL RESULTS

1) CONCENTRATION OF PM_{10} IN KALLIO STATION

Table 3 shows the forecasting results for the PM_{10} concentration time series (Kallio Station) using single and combination models. Table 4 presents a ranking regarding each metric result. The value column Mean presents an average of the positions achieved regarding all metrics, while the column Rank order the values of the means. We followed the rank to define the best prediction models. This premise is adopted to all PM series. Note that the best error values are highlighted in bold for each approach: single and combination models.

The results summarized in Tables 3 and 4 allow some important remarks. There is no perfect correspondence between the error metrics. While the smallest ARV is related to the AR model, the best MSE, MAPE, MAE, and RMSE

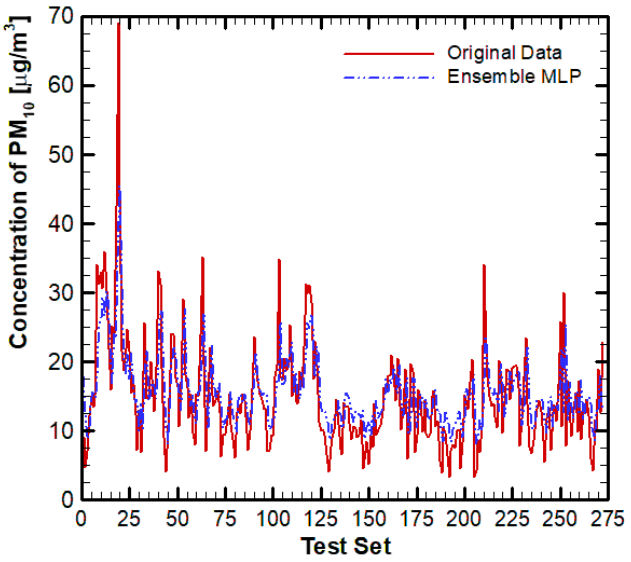

FIGURE 6. The best forecast for the test set of the PM_{10} concentration time series for Kallio station.

belong to the Ensemble with MLP. The best IA is from the single MLP.

Among the linear models, the IIR filter performed better than the AR and ARMA, considering the six metrics. The best single ANN was the MLP, which presented the best MSE for a single model. It is essential since some references [42], [57] consider the MSE as the most important error metric in time series forecasting. It is the metric minimized during the adjustment of the linear models and ANN.

Comparing the combination approaches, the MLP ensemble achieved the best performance considering four metrics, while the ELM (CR) in one (ARV), and the Median in another (IA). However, observing the general results considering the 16 predictors, the trainable ensembles stood out, reaching the best overall performances for 5 out of 6 metrics. Besides, the trainable ensembles achieved the first three positions in Table 4.

Figure 6 presents the best execution of the MLP Ensemble, the best predictor for the Kallio PM_{10} series in the test set.

2) CONCENTRATION OF PARTICULATE MATTER $PM_{2.5}$ IN KALLIO STATION

Table 5 shows the performances for $PM_{2.5}$ concentration series of Kallio Station, and Table 6 shows the ranking. Combination approaches have reached the best values for all evaluation measures. Among the linear models, the IIR filter was the best again. It is observed that the AR achieved the smallest ARV. The ELM was the best single approach.

Considering the combination models, the MLP ensemble was the one that achieved the best performance for MSE, MAPE, MAE, and RMSE. The single LR (FS) combiner achieved the best result in terms of IA.

The ranking in Table 6 shows a draw between the ELM Ensemble and MLP Ensemble, although the last achieved the best MSE. The third place was occupied by the LR (FS) combiner. Figure 7 shows the test set of the actual series and the forecasting ones.

TABLE 5. Evaluation metrics for forecasting $PM_{2.5}$ series (Kallio Station).

	Model	MSE	MAPE	ARV	IA	MAE	RMSE
Single	AR	12.29	37.81	**1.10**	0.76	2.48	3.50
	ARMA	11.47	38.35	1.55	0.78	2.43	3.39
	IIR Filter	11.46	38.09	1.53	0.78	**2.41**	3.38
	RBF	11.53	38.76	1.38	0.77	2.47	3.39
	MLP	11.49	37.90	1.67	0.78	2.42	3.39
	ELM	**11.01**	**37.80**	1.31	**0.78**	2.44	**3.32**
	ESN	12.09	40.31	1.69	0.77	2.50	3.48
	ANFIS	11.17	38.00	1.40	0.78	2.44	3.34
Combination	Mean	11.27	37.87	1.44	0.78	2.41	3.36
	Median	11.49	38.08	1.51	0.78	2.42	3.39
	LR	11.04	36.80	1.33	0.78	2.42	3.32
	LR (FS)	10.84	36.57	1.39	**0.78**	2.39	3.29
	ELM	10.75	36.54	**1.23**	0.76	2.37	3.28
	ELM (CR)	10.76	36.07	1.27	0.76	2.37	3.28
	MLP	**10.43**	**32.49**	1.30	0.75	**2.32**	**3.23**
	Copulas	10.94	37.58	1.32	0.78	2.42	3.31

TABLE 6. Ranking of the single and combination models by metric $PM_{2.5}$ series of the Kallio Station.

	Model	MSE	MAPE	ARV	IA	MAE	RMSE	Mean	Rank
Single	AR	16	8	1	14	15	16	11.67	14
	ARMA	11	14	14	8	11	11	11.5	13
	IIR Filter	10	13	13	9	6	10	10.17	10
	RBF	14	15	8	11	14	14	12.67	15
	MLP	13	10	15	6	9	13	11	11
	ELM	6	7	5	2	13	6	6.5	7
	ESN	15	16	16	12	16	15	15	16
	ANFIS	8	11	10	5	12	8	9	9
Combination	Mean	9	9	11	6	5	9	8.17	8
	Median	12	12	12	10	8	12	11	11
	LR	7	5	7	4	7	7	6.17	6
	LR (FS)	4	4	9	1	4	4	4.33	3
	ELM	2	3	2	13	2	2	4	1
	ELM (CR)	3	2	3	15	3	3	4.83	4
	MLP	1	1	4	16	1	1	4	1
	Copulas	5	6	6	3	10	5	5.83	5

TABLE 7. Evaluation metrics for forecasting PM_{10} series (Vallila Station).

	Model	MSE	MAPE	ARV	IA	MAE	RMSE
Single	AR	56.75	**38.59**	**1.03**	0.78	5.39	7.53
	ARMA	54.79	39.69	1.58	0.79	5.34	7.40
	IIR Filter	54.16	38.93	1.37	0.79	5.30	7.36
	RBF	58.34	42.91	1.30	0.78	5.62	7.64
	MLP	55.43	40.09	1.28	0.79	5.39	7.44
	ELM	**52.47**	40.42	1.33	**0.79**	**5.27**	**7.24**
	ESN	53.08	41.25	1.35	0.79	5.41	7.29
	ANFIS	57.92	42.15	1.32	0.78	5.59	7.61
Combination	Mean	53.90	40.00	1.31	0.79	5.33	7.34
	Median	54.01	39.91	1.34	0.79	5.33	7.35
	LR	53.80	40.14	1.19	0.79	5.34	7.33
	LR (FS)	51.91	39.01	1.20	0.79	5.21	7.20
	ELM	51.65	39.57	1.15	0.79	5.20	7.19
	ELM (CR)	52.95	39.46	**0.99**	0.80	5.24	7.28
	MLP	**49.10**	**31.28**	1.06	**0.80**	**5.01**	**7.01**
	Copulas	52.26	40.01	1.36	0.79	5.23	7.23

3) CONCENTRATION OF PM_{10} IN VALLILA STATION

Table 7 shows the computational results of the single and combined models for Vallila station PM_{10} concentration, and Table 8 the general ranking.

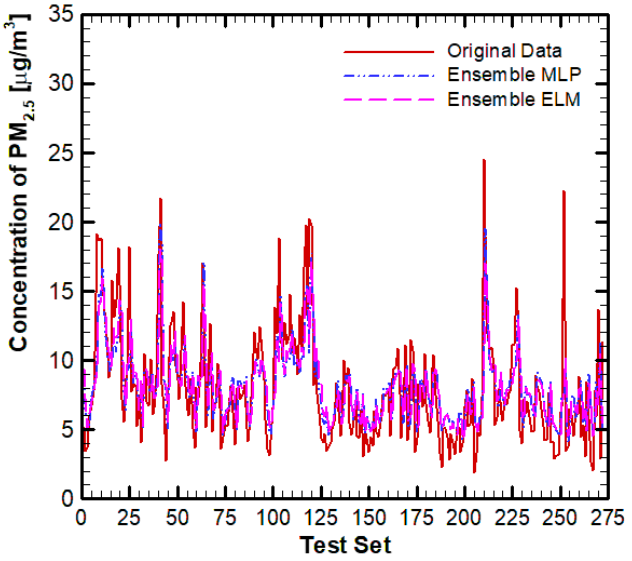

FIGURE 7. Best forecast for the test set of the $PM_{2.5}$ concentration time series for Kallio station.

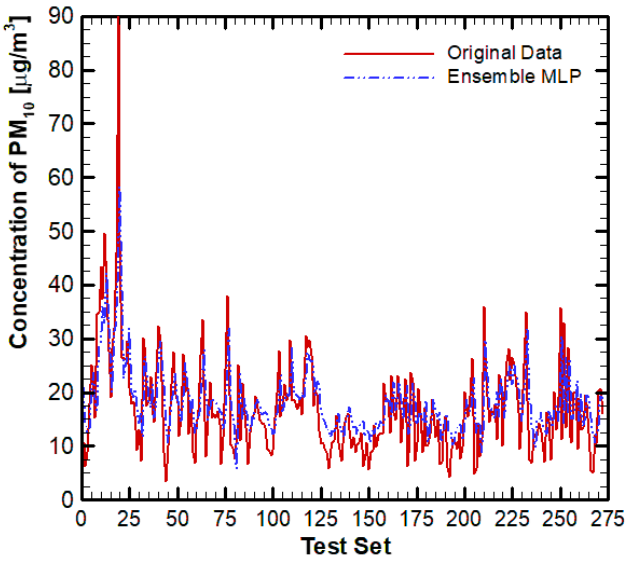

FIGURE 8. Best forecast for the test set of the PM_{10} concentration time series for Vallila station.

TABLE 8. Ranking of the single and combination models by metric for PM_{10} series of the Vallila Station.

	Model	MSE	MAPE	ARV	IA	MAE	RMSE	Mean	Rank
Single	AR	14	2	2	15	13	14	10	11
	ARMA	12	7	16	8	11	12	11	13
	IIR Filter	11	3	15	9	7	11	9.33	8
	RBF	16	16	8	14	16	16	14.33	16
	MLP	13	11	7	13	12	13	11.5	14
	ELM	5	13	11	5	6	5	7.5	6
	ESN	7	14	13	7	14	7	10.33	12
	ANFIS	15	15	10	15	15	15	14.17	15
Combination	Mean	9	9	9	12	9	9	9.5	9
	Median	10	8	12	10	8	10	9.67	10
	LR	8	12	5	5	10	8	8	7
	LR (FS)	3	4	6	4	3	3	3.83	2
	ELM	2	6	4	10	2	2	4.33	4
	ELM (CR)	6	5	1	2	5	6	4.17	3
	MLP	1	1	3	1	1	1	1.33	1
	Copulas	4	10	14	3	4	4	6.5	5

Analyzing the results in Tables 7 and 8, it is clear that the IIR Filter was the best linear model, even though, the AR reached good values for MAPE and ARV. Among the single models, the ELM was the best for four error metrics.

Considering the ensembles, the MLP combiner was the best for 5 out of 6 metrics, and the ELM (CR) in one metric, precisely the ARV. Also, the first five ranked methods belong to the combination class.

Figure 8 presents the prediction models' general behavior, showing the test set of the actual values for PM_{10} and the predictions of the best predictior (MLP Ensemble).

4) CONCENTRATION OF $PM_{2.5}$ IN VALLILA STATION

Tables 9 and 10 show the performance for $PM_{2.5}$ concentration of Vallila Station, and the ranking of the performances, respectively.

The analysis of Table 9 reveals a distinct behavior regarding the linear models. Considering all single models, the

TABLE 9. Evaluation metrics for forecasting $PM_{2.5}$ series (Vallila Station).

	Model	MSE	MAPE	ARV	IA	MAE	RMSE
Single	AR	13.40	**37.17**	**1.55**	0.74	2.67	3.66
	ARMA	12.50	38.33	2.18	**0.75**	**2.63**	3.54
	IIR Filter	12.64	38.94	2.11	0.75	2.64	3.55
	RBF	13.06	39.21	1.57	0.74	2.70	3.61
	MLP	12.81	38.93	1.87	0.75	2.65	3.58
	ELM	**12.35**	38.93	1.73	0.75	2.66	**3.51**
	ESN	13.01	41.66	2.31	0.75	2.74	3.61
	ANFIS	12.86	38.63	1.62	0.75	2.68	3.59
Combination	Mean	12.53	38.56	1.86	0.75	2.64	3.54
	Median	12.54	38.65	1.91	0.75	2.63	3.54
	LR	12.19	38.08	1.68	**0.76**	2.63	3.49
	LR (FS)	12.67	38.83	1.53	0.75	2.69	3.56
	ELM	12.07	38.33	1.60	0.72	2.64	3.47
	ELM (CR)	11.67	37.44	1.68	0.73	2.56	3.42
	MLP	**11.66**	**34.17**	**1.49**	0.72	**2.55**	**3.41**
	Copulas	12.62	39.26	1.55	0.75	2.70	3.55

TABLE 10. Ranking of the single and combination models by metric for $PM_{2.5}$ series of the Vallila Station.

	Model	MSE	MAPE	ARV	IA	MAE	RMSE	Mean	Rank
Single	AR	16	2	4	13	11	16	10.33	12
	ARMA	6	6	15	2	3	6	6.33	4
	IIR Filter	10	13	14	5	8	10	10	11
	RBF	15	14	5	12	14	15	12.5	15
	MLP	12	12	12	10	9	12	11.17	14
	ELM	5	11	10	3	10	5	7.33	7
	ESN	14	16	16	7	16	14	13.83	16
	ANFIS	13	8	7	11	12	13	10.67	13
Combination	Mean	7	7	11	4	7	7	7.17	6
	Median	8	9	13	6	5	8	8.17	8
	LR	4	4	8	1	4	4	4.17	2
	LR (FS)	11	10	2	9	13	11	9.33	9
	ELM	3	5	6	16	6	3	6.5	5
	ELM (CR)	2	3	9	14	2	2	5.33	3
	MLP	1	1	1	15	1	1	3.33	1
	Copulas	9	15	3	8	15	9	9.83	10

ARMA achieved the best IA and MAE errors, while AR the smallest MAPE and ARV. Table 10 shows that ARMA was

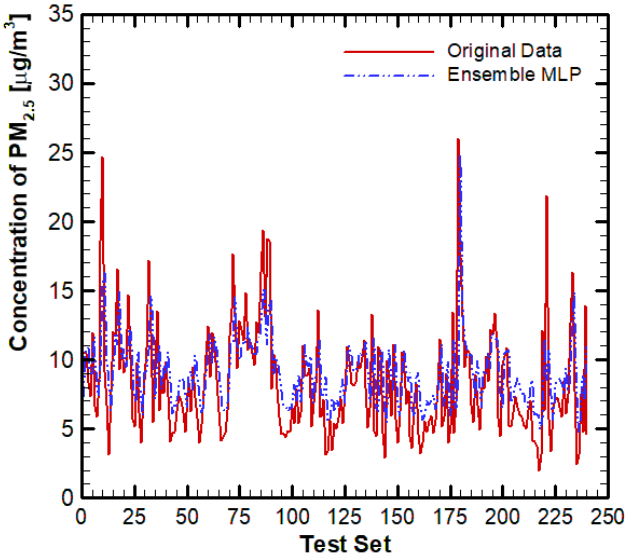

FIGURE 9. The best forecast for the test set of the $PM_{2.5}$ concentration time series for Vallila station.

TABLE 11. Evaluation metrics for forecasting PM_{10} series (São Paulo Station).

	Model	MSE	MAPE	ARV	IA	MAE	RMSE
Single	AR	171.15	42.95	1.00	0.80	9.85	13.08
	ARMA	163.77	40.37	1.04	0.81	9.75	12.80
	IIR Filter	168.69	38.95	0.96	0.80	9.72	12.99
	RBF	179.75	40.17	1.09	0.80	10.16	13.41
	MLP	165.12	37.76	0.92	0.81	9.65	12.85
	ELM	**156.16**	**37.16**	0.91	**0.81**	**9.40**	**12.50**
	ESN	164.93	39.11	0.89	0.81	9.68	12.84
	ANFIS	170.92	37.17	**0.88**	0.80	10.09	13.07
Combination	Mean	154.33	37.81	0.96	0.82	9.39	12.42
	Median	158.61	38.16	0.97	0.81	9.46	12.59
	LR	152.82	36.85	0.73	0.82	9.61	12.36
	LR (FS)	153.29	34.99	0.78	0.82	9.53	12.38
	ELM	**141.87**	34.33	0.63	**0.86**	**9.22**	**11.91**
	ELM (CR)	145.89	35.52	0.67	0.86	9.36	12.08
	MLP	145.09	**31.33**	**0.63**	0.86	9.24	12.04
	Copulas	149.59	36.83	0.79	0.82	9.52	12.23

the fourth-best predictor for Vallila $PM_{2.5}$ series. Considering the six metrics, the IIR filter presents a better raking position than AR.

On the other hand, the ELM obtained the smallest MSE and RMSE regarding the single predictors. About the combination models, the MLP Ensemble was also the winner, followed by the LR combiner and ELM (CR) proposal. Note that, in general, the combination models stood out.

Figure 9 shows the test set of the actual values for $PM_{2.5}$ and the MLP combination model's forecasts.

5) CONCENTRATION OF PM_{10} IN SÃO PAULO STATION

Table 11 summarizes the computational results found by 16 forecasting models for PM_{10} concentration in São Paulo, while Table 12 presents the performances' ranking.

For São Paulo PM_{10} concentration, the computational performances were uniform. The linear approaches overcame

TABLE 12. Ranking of the single and combination models by metric for PM_{10} series of the São Paulo Station.

	Model	MSE	MAPE	ARV	IA	MAE	RMSE	Mean	Rank
Single	AR	15	16	14	15	14	15	14.83	15
	ARMA	10	15	15	10	13	10	12.17	13
	IIR Filter	13	12	12	13	12	13	12.5	14
	RBF	16	14	16	16	16	16	15.67	16
	MLP	12	9	10	11	10	12	10.67	10
	ELM	8	7	9	8	5	8	7.5	7
	ESN	11	13	8	12	11	11	11	11
	ANFIS	14	8	7	14	15	14	12	12
Combination	Mean	7	10	11	7	4	7	7.67	8
	Median	9	11	13	9	6	9	9.5	9
	LR	5	6	4	6	9	5	5.83	6
	LR (FS)	6	3	5	5	8	6	5.5	5
	ELM	1	2	2	1	1	1	1.33	1
	ELM (CR)	3	4	3	3	3	3	3.17	3
	MLP	2	1	1	2	2	2	1.67	2
	Copulas	4	5	6	4	7	4	5	4

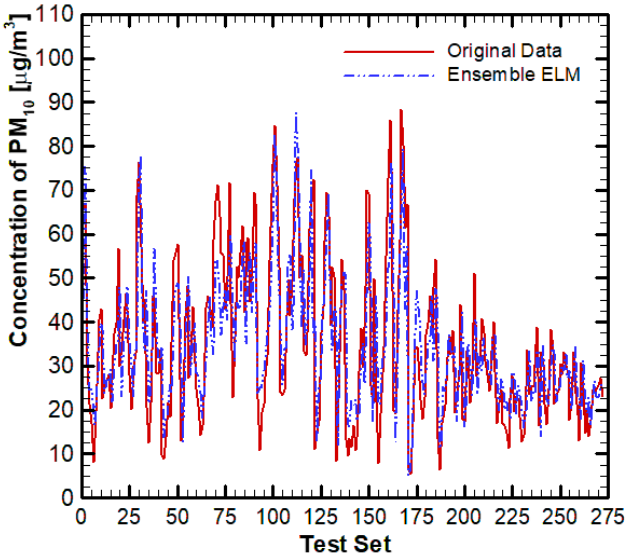

FIGURE 10. The best forecast for the test set of the PM_{10} concentration time series for São Paulo station.

just the RBF, being the ARMA the best of them. The ELM overcame some ensembles, being the best single model regarding five metrics (except ARV).

The six best methods belong to the ensembles based on feedforward neural models, MLP and ELM. However, different from previous cases, the ELM combiner was the best (winner for four metrics), followed by the MLP and ELM (CR).

Figure 10 sketches the forecasting of the ELM ensemble.

6) CONCENTRATION OF $PM_{2.5}$ IN SÃO PAULO STATION

We present in this section the computational results for São Paulo $PM_{2.5}$ in Table 13, together with the ranking in Table 14.

For São Paulo $PM_{2.5}$ concentration time series, the performances showed an remarkable behavior. The ELM was the winner for all metrics regarding the single models, while

TABLE 13. Evaluation metrics for forecasting $PM_{2.5}$ series (São Paulo Station).

	Model	MSE	MAPE	ARV	IA	MAE	RMSE
Single	AR	80.46	43.60	0.95	0.79	6.56	8.97
	ARMA	79.74	41.14	1.06	0.80	6.51	8.93
	IIR Filter	79.17	41.76	1.14	0.80	6.52	8.90
	RBF	85.04	42.13	1.05	0.78	6.71	9.22
	MLP	78.05	41.74	1.03	0.80	6.48	8.83
	ELM	**76.79**	**40.08**	**0.93**	**0.80**	**6.43**	**8.76**
	ESN	79.28	42.84	0.96	0.79	6.65	8.90
	ANFIS	84.00	41.39	1.06	0.79	6.66	9.16
Combination	Mean	77.26	40.98	1.02	0.80	6.40	8.79
	Median	76.96	40.83	1.03	0.80	6.40	8.77
	LR	79.53	43.01	1.13	0.80	6.57	8.92
	LR (FS)	76.24	40.50	1.00	0.80	6.42	8.73
	ELM	76.85	40.08	0.96	0.81	6.39	8.77
	ELM (CR)	76.64	41.21	1.02	0.81	6.43	8.75
	MLP	**74.11**	**33.30**	**0.82**	**0.82**	**6.28**	**8.61**
	Copulas	78.64	42.75	1.01	0.80	6.52	8.87

TABLE 14. Ranking of the single and combination models by metric for $PM_{2.5}$ series of the São Paulo Station.

	Model	MSE	MAPE	ARV	IA	MAE	RMSE	Mean	Rank
Single	AR	14	16	3	14	12	14	12.17	13
	ARMA	13	7	14	10	9	13	11	11
	IIR Filter	10	11	16	8	10	10	10.83	10
	RBF	16	12	12	16	16	16	14.67	16
	MLP	8	10	11	9	8	8	9	8
	ELM	4	2	2	6	7	4	4.17	4
	ESN	11	14	4	13	14	11	11.17	12
	ANFIS	15	9	13	15	15	15	13.67	15
Combination	Mean	7	6	8	7	4	7	6.5	7
	Median	6	5	10	5	3	6	5.83	6
	LR	12	15	15	12	13	12	13.17	14
	LR (FS)	2	4	6	4	5	2	3.83	3
	ELM	5	3	5	2	2	5	3.67	2
	ELM (CR)	3	8	9	3	6	3	5.33	5
	MLP	1	1	1	1	1	1	1	1
	Copulas	9	13	7	10	11	9	9.83	9

the MLP Ensemble was the general winner considering all models. The ELM achieved the fourth-best performance. The best models were, again, the MLP and ELM. Interestingly, the Ensemble LR (FS) was the third best, but the LS was one of the worst methods. Concerning the linear models, the IIR Filter and the ARMA presented a similar Mean value.

Figure 11 shows the time behavior of São Paulo $PM_{2.5}$ for the test set in comparison to the original data.

7) CONCENTRATION OF PM_{10} IN CAMPINAS STATION

In Tables 15 and 16 we show the general performances for Campinas PM_{10} prediction.

In Campinas' case, the ARMA and IIR Filter present almost a draw considering the final ranking score, with a small advantage for the ARMA. Both overcame the AR. Also, this is the first time we see the MLP present a Mean worse than the linear approaches, despite it reached the best IA.

We observed the ELM as the second general best model, being the general best for MAE, and the winner for 3 metrics considering just the single models. Also, the ESN was highlighted regarding the single models for MSE and RMSE. Considering all predictors, the ESN was the sixth-best.

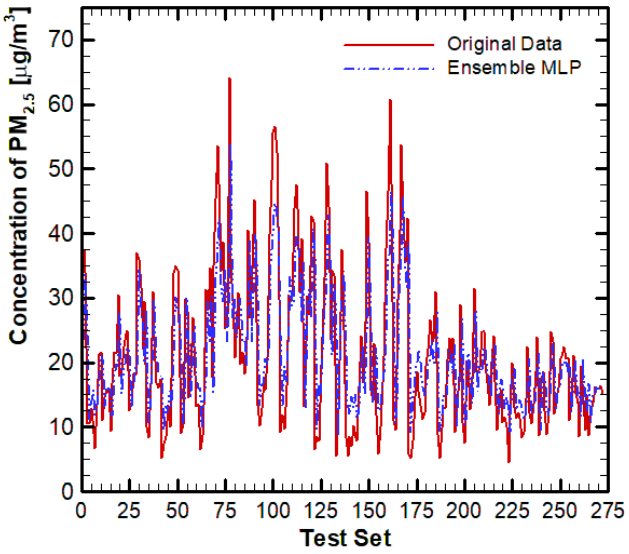

FIGURE 11. Best forecast for the test set of the $PM_{2.5}$ concentration time series for São Paulo station.

TABLE 15. Evaluation metrics for forecasting PM_{10} series (Campinas Station).

	Model	MSE	MAPE	ARV	IA	MAE	RMSE
Single	AR	96.48	23.47	0.83	0.78	7.30	9.82
	ARMA	92.62	23.68	1.10	0.79	7.36	9.62
	IIR Filter	92.92	23.50	1.15	0.79	7.37	9.64
	RBF	100.60	23.63	0.76	0.77	7.52	10.03
	MLP	90.57	23.91	1.20	**0.80**	7.48	9.52
	ELM	89.20	**22.23**	**0.72**	0.78	**7.01**	9.44
	ESN	**88.69**	22.81	0.84	0.79	7.19	**9.42**
	ANFIS	98.67	23.29	0.77	0.77	7.39	9.93
Combination	Mean	**85.95**	22.54	0.88	0.79	**7.04**	**9.27**
	Median	89.45	22.60	0.93	0.79	7.12	9.46
	LR	98.16	23.38	0.75	0.77	7.37	9.91
	LR (FS)	99.15	23.56	0.73	0.77	7.42	9.96
	ELM	93.14	22.60	0.73	0.83	7.17	9.65
	ELM (CR)	93.61	22.53	0.73	0.83	7.15	9.67
	MLP	86.40	**21.85**	0.74	**0.83**	7.05	9.29
	Copulas	98.44	23.20	**0.72**	0.77	7.39	9.92

TABLE 16. Ranking of the single and combination models by metric for PM_{10} series of the Campinas Station.

	Model	MSE	MAPE	ARV	IA	MAE	RMSE	Mean	Rank
Single	AR	11	11	10	11	8	11	10.33	11
	ARMA	7	15	14	8	9	7	10	8
	IIR Filter	8	12	15	7	11	8	10.17	9
	RBF	16	14	8	16	16	16	14.33	16
	MLP	6	16	16	4	15	6	10.5	12
	ELM	4	2	2	10	1	4	3.83	2
	ESN	3	7	11	6	7	3	6.17	6
	ANFIS	14	9	9	12	13	14	11.83	14
Combination	Mean	1	4	12	5	2	1	4.17	3
	Median	5	6	13	8	4	5	6.83	7
	LR	12	10	7	13	10	12	10.67	13
	LR (FS)	15	13	5	15	14	15	12.83	15
	ELM	9	5	3	3	6	9	5.83	5
	ELM (CR)	10	3	4	2	5	10	5.67	4
	MLP	2	1	6	1	3	2	2.5	1
	Copulas	13	8	1	14	12	13	10.17	9

The ensembles followed the same tendency as the previous simulations, being the MLP combiner the winner. However,

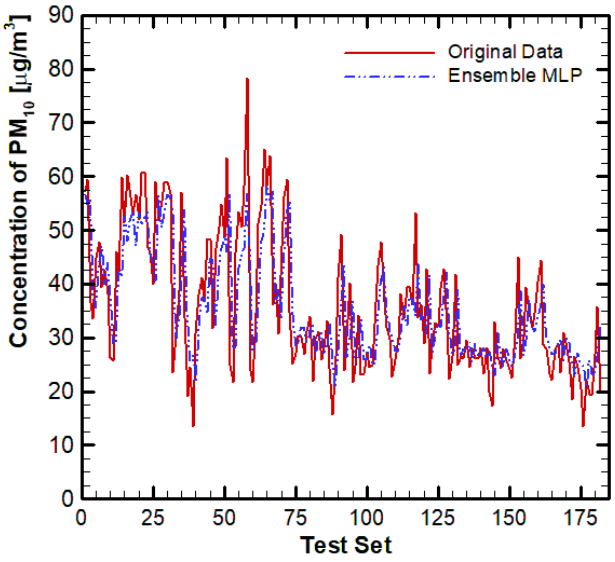

FIGURE 12. The best forecast for the test set of the PM_{10} concentration time series for Campinas station.

TABLE 17. Evaluation metrics for forecasting PM_{10} series (Ipojuca Station).

	Model	MSE	MAPE	ARV	IA	MAE	RMSE
Single	AR	36.51	17.78	1.18	0.80	4.77	6.04
	ARMA	**34.05**	**17.29**	1.29	**0.81**	**4.63**	**5.84**
	IIR Filter	35.27	17.66	1.37	0.81	4.68	5.94
	RBF	35.91	17.85	1.15	0.80	4.78	5.99
	MLP	35.06	17.50	1.16	0.80	4.75	5.92
	ELM	36.01	17.67	**1.14**	0.80	4.75	6.00
	ESN	36.12	17.99	1.27	0.80	4.76	6.01
	ANFIS	37.51	17.89	1.27	0.80	4.84	6.12
Combination	Mean	34.02	17.32	1.24	0.81	4.62	5.83
	Median	33.77	17.22	1.20	0.81	4.59	5.81
	LR	33.46	16.56	0.92	0.80	4.60	5.78
	LR (FS)	33.30	16.53	0.93	0.81	4.58	5.77
	ELM	33.43	16.65	0.91	0.83	4.59	5.78
	ELM (CR)	32.94	16.22	0.85	**0.84**	4.56	5.74
	MLP	**29.56**	**13.95**	**0.81**	0.83	**4.37**	**5.44**
	Copulas	36.21	17.91	1.20	0.80	4.75	6.02

the ensemble using the Mean achieved the third position in the ranking, reaching the best MSE and RMSE, and the second-best MAE. The ELM-based approaches were good again. On the other hand, the LR and LR (FS) ensembles appeared between the worst predictors.

Figure 12 shows the graphic of the MLP Ensemble and the observed data for the Campinas time series.

8) CONCENTRATION OF PM_{10} IN IPOJUCA STATION

Finally, we present in Tables 17 and 18 the general performance for the last series, the Ipojuca PM_{10} concentration, regarding the error values and the ranking of the forecasting models, respectively.

The results show that the ARMA was the best single model, being the winner for five metrics when considering just the single models. The MLP was the second-best in this group.

TABLE 18. Ranking of the single and combination models by metric for PM_{10} series of the Ipojuca Station.

	Model	MSE	MAPE	ARV	IA	MAE	RMSE	Mean	Rank
Single	AR	15	12	9	15	14	15	13.33	14
	ARMA	8	7	15	4	8	8	8.33	8
	IIR Filter	10	10	16	8	9	10	10.5	10
	RBF	11	13	7	12	15	11	11.5	12
	MLP	9	9	8	10	11	9	9.33	9
	ELM	12	11	6	14	12	12	11.17	11
	ESN	13	16	14	11	13	13	13.33	14
	ANFIS	16	14	13	16	16	16	15.17	16
Combination	Mean	7	8	12	6	7	7	7.83	7
	Median	6	6	10	5	4	6	6.17	6
	LR	5	4	4	9	6	5	5.5	5
	LR (FS)	3	3	5	7	3	3	4	3
	ELM	4	5	3	3	5	4	4	3
	ELM (CR)	2	2	2	1	2	2	1.83	2
	MLP	1	1	1	2	1	1	1.17	1
	Copulas	14	15	11	13	10	14	12.83	13

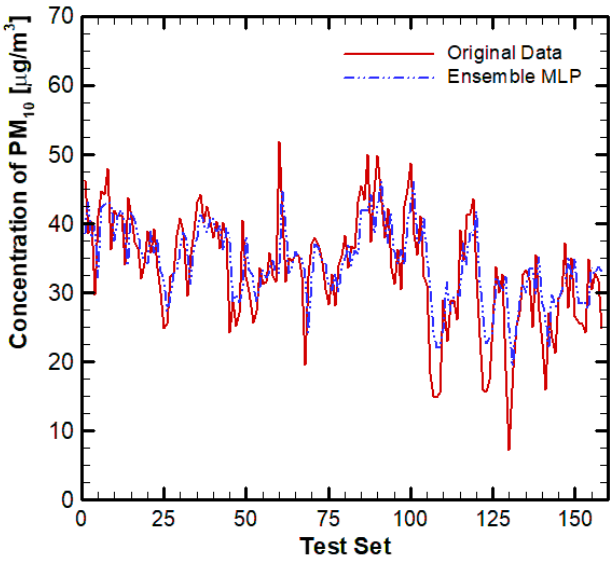

FIGURE 13. The best forecast for the test set of the PM_{10} concentration time series for Ipojuca station.

We highlight that the ELM presented its worst position in the ranking.

The combination models in the Ipojuca case stood out. Just the Copulas combiner was worse than most of the single models. The MLP Ensemble won for five metrics (the best one), and the ELM (CR) for IA (the second general best). The next was the ELM and the LR (FS) combiners.

Figure 13 shows the output of the MLP Ensemble for the Ipojuca PM_{10} time series.

IV. DISCUSSION

After the presentation and initial discussion of the results regarding the 16 forecasting models and 8 time series, we show Table 20, which summarizes the general prediction performance, with the following correspondence: K10 - Kallio PM_{10}; K2.5 - Kallio $PM_{2.5}$; V10 - Vallila PM_{10}; V2.5 - Vallila $PM_{2.5}$; SP10 - São Paulo PM_{10}; SP2.5 - São

TABLE 19. The general ranking of performances considering the eight time series taken into account.

	Series	K10	K2.5	V10	V2.5	SP10	SP2.5	C10	I10	Mean	Rank
Single	AR	13	14	11	12	15	13	11	14	12.88	14
	ARMA	12	13	13	4	13	11	8	8	10.25	11
	IIR Filter	4	10	8	11	14	10	9	10	9.50	10
	RBF	15	15	16	15	16	16	16	12	15.13	16
	MLP	4	11	14	14	10	8	12	9	10.25	11
	ELM	7	7	6	7	7	4	2	11	6.38	5
	ESN	10	16	12	16	11	12	6	14	12.13	13
	ANFIS	16	9	15	13	12	15	14	16	13.75	15
Combination	Mean	4	8	9	6	8	7	3	7	6.50	6
	Median	7	11	10	8	9	6	7	6	8.00	7
	LR	14	6	7	2	6	14	13	5	8.38	9
	LR (FS)	10	3	2	9	5	3	15	3	6.25	4
	ELM	2	1	4	5	1	2	5	3	2.88	2
	ELM (CR)	3	4	3	3	3	5	4	2	3.38	3
	MLP	1	1	1	1	2	1	1	1	1.13	1
	Copulas	9	5	5	10	4	9	9	13	8.00	7

TABLE 20. Number of studentized residuals greater than 3.0 in absolute value (Kallio -K10, K2.5, Vallila - V10, V2.5, Campinas - C10, Ipojuca - I10, and São Paulo - SP10, SP2.5). The worst performance per series is in bold.

	Series	K10	K2.5	V10	V2.5	SP10	SP2.5	C10	I10
Single	AR	15	17	16	14	**15**	14	11	8
	ARMA	15	**20**	17	11	12	13	11	7
	IIR Filter	15	19	18	12	9	13	11	9
	RBF	13	18	16	13	11	12	8	7
	MLP	14	**20**	17	11	13	14	**12**	7
	ELM	15	13	16	11	12	13	8	7
	ESN	**20**	16	**19**	**16**	14	10	10	**10**
	ANFIS	15	19	16	13	10	11	7	6
Combination	Mean	15	15	17	12	11	12	9	8
	Median	15	19	18	12	11	13	10	9
	LR	17	18	15	12	10	**16**	9	8
	LR (FS)	15	16	17	12	10	11	9	9
	ELM	16	16	18	12	10	12	9	8
	ELM (CR)	16	17	**19**	12	9	12	9	8
	MLP	16	16	16	11	10	12	10	8
	Copulas	16	16	**19**	12	8	13	9	7

Paulo $PM_{2.5}$; C10 - Campinas PM_{10}; I10 - Ipojuca PM_{10}; Mean - the mean ranking from all time series to each applied model; and Rank - the new mean-based ranking.

From the results in Tables 3 to 20, one can discuss many aspects. For the linear models, ARMA and IIR Filter achieved the best results four times. It seems to be clear that recursive linear models (ARMA and IIR Filter) have an advantage over AR, which does not present feedback recurrences. Despite the absence of closed-form solutions to determine the free parameters, the inclusion of feedback information to perform the models output response seems to compensate this drawback.

Although we can not state which linear approach is more suitable, it is essential to highlight that the IIR filter is mainly applied to adaptive filtering problems [41], while the ARMA in forecasting tasks. Therefore, if the user chooses to use linear approaches, the IIR filter should be considered. Also, a further investigation on the use of bio-inspired metaheuristics should be conduced, analyzing other approaches, such as Genetic Algorithms, Differential Evolution, and so on [93].

In general, the linear models did not overcome the ensembles, but they perform better than some nonlinear approaches. Besides, the response of the models may increase the diversity to perform the ensembles.

About the single ANN-models, it was not expected that RBF, ANFIS, and ESN were worse than the linear ones. Among them, the ESN was the best predictor. Some previous studies revealed the prediction capability of such approaches in related time series forecasting [51], [94]. However, as in the linear case, these models must have generated diversity in the final response.

The application of the MLP led to intriguing observations. When used as a combiner, the architecture led to the best performances. However, considering the general ranking, it presented similar performance to ARMA and IIR Filter. But, for the Vallila series, it shows relatively poor performance, figuring among the worse predictors.

Undoubtedly, the best single approach was ELM. Except for the Ipojuca series, the model was always among the first half of the ranked methods. For Campinas, it was the second-best. It is an important remark since the ELM is similar to the traditional MLP, but as its hidden neurons are not tuned, the computational training effort is relatively lower. In addition, these results reinforce the premises found in other time series forecasting problems [51], [57].

The computational results were favorable to the use of combination models (ensembles). Table 20 reveals, in 7 out of 8 cases, the MLP ensemble reached the best raking position, followed by the two ELM ensembles. It is intriguing since, in the single approaches, the ELM performed better. Due to the advantage in terms of performance, we can state that the trainable ensembles, mainly those endowed by feedforward neural models, are more suitable to solve the prediction task.

The use of the CR by the ELM Ensemble does not prove to be an advantage. We observed performance gains in two cases. However, as the computational cost involved in its application is not high it may be important for other series.

The non-trainable approaches were similar in terms of performance, with an advantage to the mean. However, they did not overcome the ELM and MLP.

Concerning the use of the linear regression (LR), we noted that the implementation of the Feature Selection (FS) process increased the quality of the results for 6 cases. Without this technique, the LR combiner was worse than the non-trainable ensembles. Also, the LR (FS) was the fourth-best model. This is important because further investigation can be conducted to evaluate FS methods to other ensembles.

The Normal Copula-based combination presented interesting results in literature [29], [80], [81]. In this investigation, it performed fairly but the it did not overcome the neural ensembles.

Another critical issue is related to the dispersion of the results considering 30 independent simulations. In this case, we used the MSE achieved for the Kallio PM_{10} series to exemplify the models behavior, as show in the boxplot graphic depicted in Figure 14. The non-trainable ensembles and the methods with close-form solution for the training,

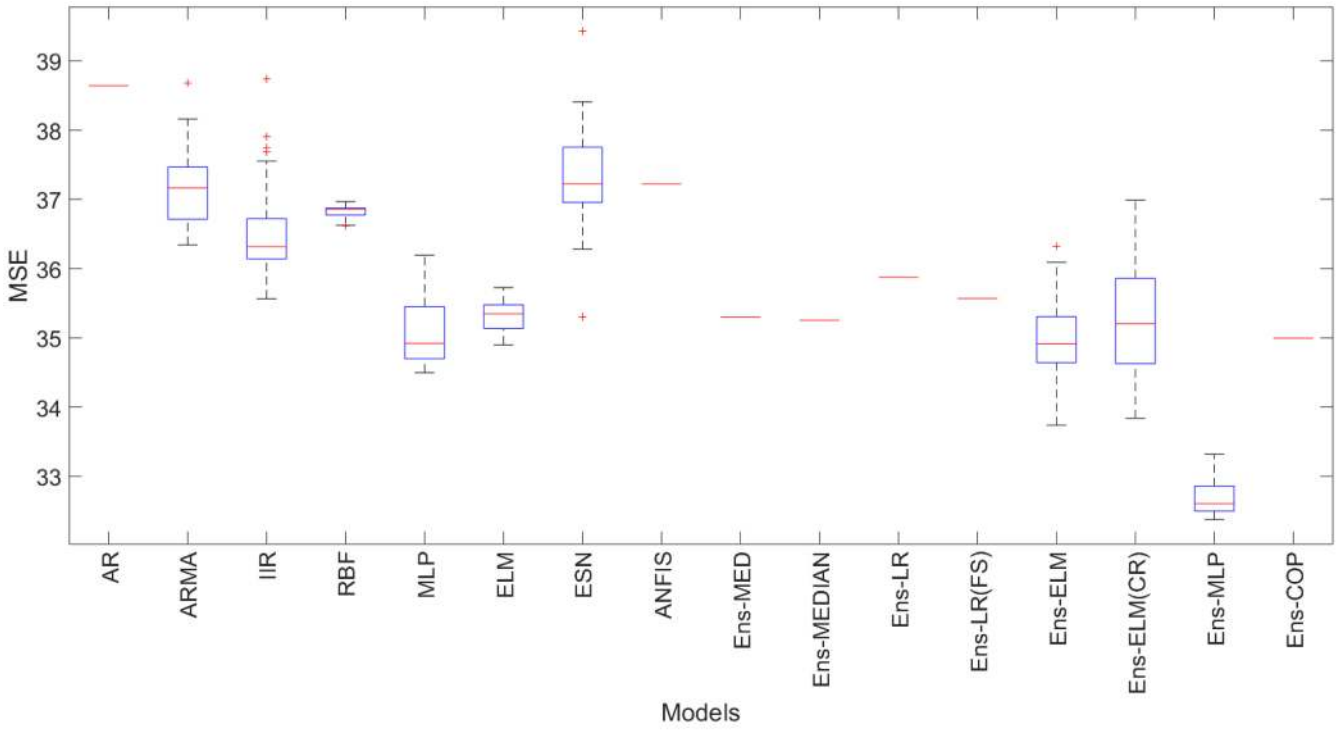

FIGURE 14. Boxplot of the MSE of the test set for Kallio PM_{10} series.

do not present dispersion, as expected. Note that the RBF presents a small dispersion, but its best performance is worse than most other single models. Also, the MLP Ensemble presented a relatively small dispersion compared to the other neural-based combiners.

Comprising the environmental aspect, we can state that the best forecasting models (Figures 6 to 13) were adequate to all considered time series, even during extreme events of high air pollution. It is a crucial behavior, as the health systems may collapse due to overcrowding during high air pollution events, which may help governments take rapid measures to ensure the safety of the whole population. But, it is important to highlight that the forecast power may vary from series to series.

A. RESIDUAL ANALYSIS

Once the target time series and respective forecasts are close in some cases, and it may be difficult to distinguish in the time series plots, the Studentized residuals were computed by using `rstudent` function of `R` software [95]. The Studentized residual measures how many standard deviations of each observed value of a time series deviates from an adjusted model considering all samples except that observation [96]. Due to the expressive number of time series and models, we present in Table 20 the number of samples showing Studentized values greater than 3.0 in absolute value to some of the studied time series.

The worst results are in bold. One can see that Ipojuca PM_{10} (I10) and Kallio $PM_{2.5}$ (K2.5) series have presented,

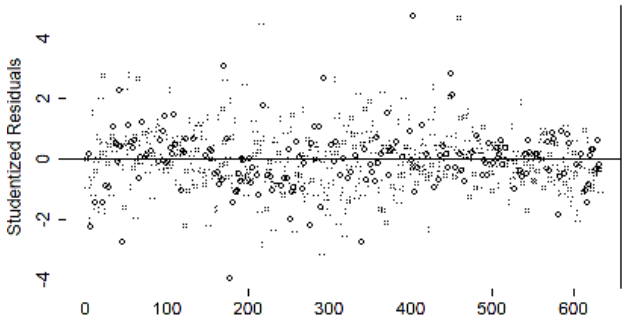

FIGURE 15. Studentized Residuals for Ipojuca PM_{10} series using ANFIS model.

in this order, the best and worst results, taking the maximum number of Studentized-based outliers into account.

Figures 15 and 16 show the Studentized residuals plot with respect to the best model (ANFIS for I10, with 6 outliers) and one of the worst predictors (single MLP for K2.5 series, with 20 outliers).

One can observe that the difference between the number of Studentized-based outliers per series was lesser than 8. These results showed no significant findings regarding the best performances. Then, the conclusion obtained with the previous error metrics analyses might be maintained.

B. LIMITATIONS AND POSSIBLE DIRECTIONS

This work intended to determine the most suitable model to perform PM forecasting, considering only endogenous

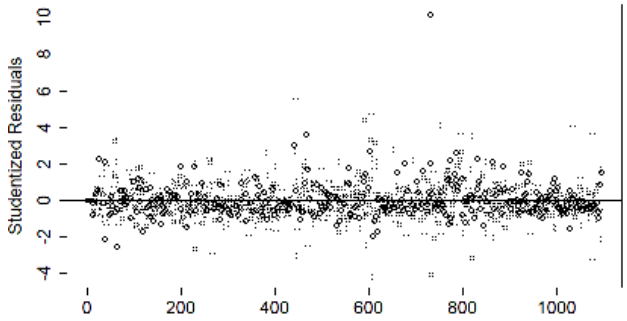

FIGURE 16. Studentized Residuals for Kallio $PM_{2.5}$ series using MLP model.

inputs (lags). A comparison of 16 approaches was carried out. A variety of techniques was considered, including linear models, neural networks, and combination models. The literature presents other proposals that use the same premise for time series forecasting [18], [57], [60], [97]. However, it is important to mention the limitations of this study, since other investigations followed different directions.

It is known that a general model for predicting the concentrations of PM, especially the fine fraction, is quite challenging to be achieved without considering source apportionment, seasonality, local topography, and climatic factors. Previous investigations proposed hybrid techniques that employ many other models, such as ANNs and SVM. Also, air trajectory models have been developed [98].

An eventual poor performance by solely using an algorithm directly on a time series should be a consequence of not including variability of local conditions. Therefore, a specific algorithm may perform better in other locations and other seasons. In this sense, a further investigation addressing calibration and sensitivity analysis considering seasonality and climatic variations is needed.

Regarding the activation functions of the neural models, we only use hyperbolic tangent (tanh) and linear function. It is a usual choice because it is fully differentiable and ranges from -1 to $+1$. However, the literature presents many possible functions, such as threshold, sigmoid, or ReLU. A change in the activation function may lead to an increase in the models performance.

The computational results were convergent on showing the ensembles behave better in PM forecasting, considering the databases addressed. Regarding the separation of the samples in training, validation, and test, we considered 50%, 25%, and 25%, respectively. This division is based on the premise that the training set must contain the temporal patterns to allow the ANN to capture the statistical oscillations of the target series over time. Note that, only in the Ipojuca series, we did not use a full year of samples in training, but something close to that. Also, we would like to provide a significant amount of samples in the test set, to obtain a better evaluation of the results. However, some studies indicate that distinct divisions can be more adequate, such as 70%, 15% and 15% [42].

TABLE 21. Weights of the single models in the NC combination for each time series taken into account (Kallio -K10, K2.5, Vallila - V10, V2.5, Campinas - CPM10, Ipojuca - GPM10, and São Paulo - SPPM10, SPPM2.5). According to the magnitude of the residuals variance and covariances, the weight of the main model in the combination is highlighted in bold.

Model	K10	K2.5	V10	V2.5	C10	G10	SP10	SP2.5
AR	0.0444	-9.4e-03	0.1998	-0.0367	0.0802	-0.2291	0.3856	0.4472
ARMA	-0.1250	-0.1958	0.3488	0.3266	-0.0813	-0.5496	0.0511	0.0678
FIR	-0.1889	0.2757	0.2964	-0.5956	0.2900	0.3158	-0.0496	0.5222
RBF	-0.7775	-0.3055	0.6968	**0.9351**	**0.9260**	-0.0682	-0.8318	0.6390
MLP	0.7511	-0.1467	-0.764	-0.0153	-0.3197	0.1318	-0.1138	-0.1310
ELM	0.3242	0.6494	**0.8592**	0.6251	0.3563	**0.7549**	0.5751	0.1950
ESN	0.0838	-0.0119	-0.0364	-7e-04	-0.0222	0.6178	0.0905	0.0411
ANFIS	**0.8879**	**0.7443**	-0.6005	-0.2386	-0.2294	0.0266	**0.8928**	**-0.7813**

In order to determine the number of hidden neurons, we performed a search in a grid. It is evident that our decision increases the computational effort since we trained many ANN topologies for each neural proposal. However, the literature provides some formulas for defining the number of hidden neurons. In an embracing study provided by Madhiarasan and Deepa [99], the authors presented 13 approaches to deal with similar task. The formulas can be useful, specially when the user has a short time or little computational power to perform the predictions.

C. INSIGHTS REGARDING DIVERSITY AND PARSIMONY

Table 21 summarizes each single model performance according to the variability of its residuals and its linear correlation with the remaining single predictors, taking the training sets into account. The greater the absolute value of the weight, the better the performance and dissimilarity of a single predictor, considering the remaining models.

Naturally, depending on the single model performance in the test set, we may have a possible overfitting during the training phase. Thus, finding parsimonious models, i.e., predictors that are accurate and efficient though involving a simplified architecture (with a reduced number of parameters) is also in the kernel of time series forecasting exercises.

For the sake of illustration, one can see the remarkable performance of ANFIS when forecasting the training set of PM_{10} concentration in Kallio station (K10). However, this model has performed poorly during test, assuming one of the worst positions (see Table 4). In the current way Copulas-Based Ensemble is modeled, it is unable to handle such a problem, also leading to poor results. In fact, it has stayed in the second half of the general rank (Table 4).

Diversity also plays an important role in ensemble models. Figure 17 brings a sketch of the level of dependence between the single models when forecasting the training set of K10. ESN seems to present the most heterogeneous results in comparison with the alternative single models taken into account. On the other hand, the expressive relationship between ARMA and IIR and between IIR and ELM illustrate the challenge of promoting diversity in ensemble studies. Considering Pearson's correlations estimates between the residuals of the single models in the training phase, one has a range from 0.5441 to 0.9929 and an average of 0.8796, taking the triple (0.8369, 0.9571, 0.9296) as quartiles.

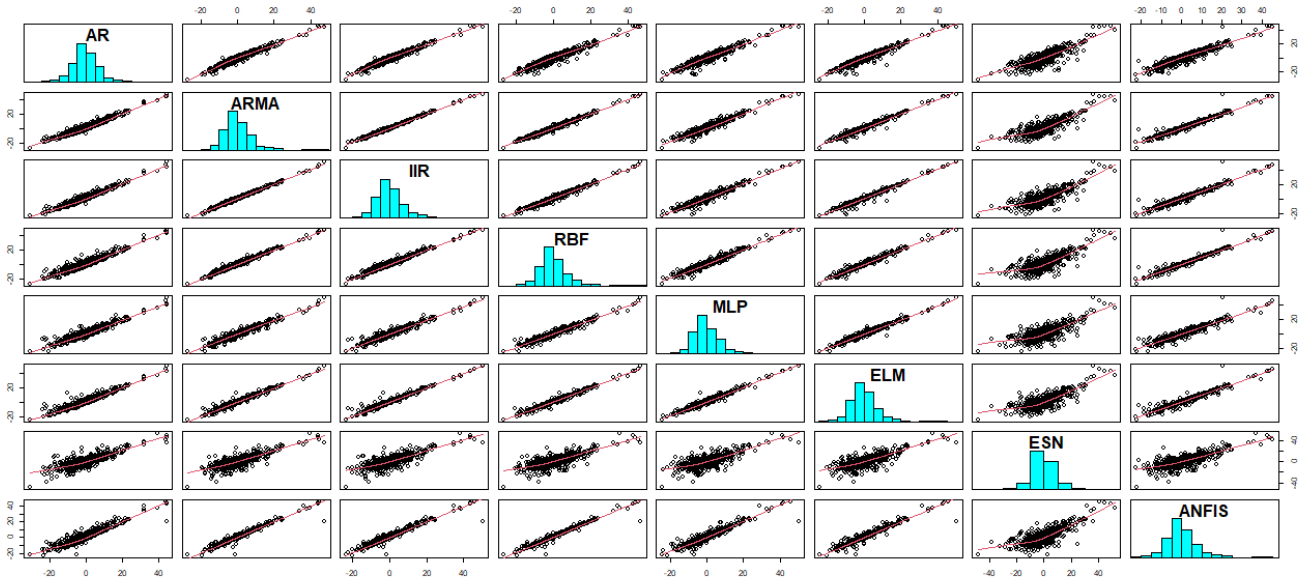

FIGURE 17. Variability and relationship between residuals of the single models when forecasting the training set of PM_{10} concentration from Kallio Station.

V. CONCLUSION

In the The present work it were evaluated single and combination methods used to particulate matter time series forecasting (PM_{10} and $PM_{2.5}$) concentration. The database comprises daily records from Kallio and Vallila stations (Finland), São Paulo, Campinas, and Ipojuca stations (Brazil).

A variety of single models was considered, including Autoregressive model (AR), Autoregressive and Moving Average Model (ARMA), Infinite Impulse Response (IIR) Filters, Multilayer Perceptron (MLP), Radial Basis Function Networks (RBF), Extreme Learning Machines (ELM), Echo State Networks (ESN), and Adaptive Network Fuzzy Inference System (ANFIS). As combination approaches (ensembles), we considered two non-data-driven combinations (median and mean), linear regression, MLP, ELM, and Normal-based Copulas.

According to a number of evaluation metrics (i.e. MSE, MAPE, ARV, IA, MAE, and RMSE), the ensemble methods led to the best overall result in all analyzed PM time series. Particularly, the MLP combination seems attractive. The capability to achieve the best results from different models is an important advantage of the ensembles. Among single models, ELM has been remarkable.

We also discussed the challenges of promoting diversity and avoiding overfitting in the single modeling phase. Ongoing research by part of the authors involves these themes. Other models can be evaluated to improve the ensemble's diversity, as Deep Neural Networks, other versions of the ESN, or hybrid systems [18].

Considering the United Nations sustainable development goals (SDG) [100], our study presents a contribution to air quality forecasting. It can advise governments to prepare hospitals during extreme air pollution events, and has premises to air pollution reduction. It is indirectly related to SDG 3-good health and well-being and SDG 11-sustainable cities and communities. Further, future works can be developed addressing pollutant series from different stations. Besides that, exogenous variables, such as weather or seasonality data, can be considered.

Copulas models that disregard from the minimal variance approach (based on the multivariate normal probability distribution) can also be studied. Finally, the proposed methodology may be tested to other air pollutants and higher forecasting horizons.

REFERENCES

[1] S. T. Ebelt, W. E. Wilson, and M. Brauer, "Exposure to ambient and nonambient components of particulate matter: A comparison of health effects," *Epidemiology*, vol. 16, no. 3, pp. 396–405, May 2005.

[2] F. Dominici, M. Greenstone, and C. R. Sunstein, "Particulate matter matters," *Science*, vol. 344, no. 6181, pp. 257–259, Apr. 2014.

[3] M. Jerrett, "The death toll from air-pollution sources," *Nature*, vol. 525, no. 7569, pp. 330–331, Sep. 2015.

[4] WHO—World Health Organization. (2002). *Reducing Risks, Promoting Healthy Life*. [Online]. Available: https://www.who.int/whr/2002/en/whr02_en.pdf?ua=1

[5] WHO—World Health Organization. (2008). *Out of 10 People World Wide Breathe Polluted Air, but More Countries are Taking Action*. [Online]. Available: http://www.who.int/news-room/detail/02-05-2018-9-out-of-10-people-worldwide-breathe-pollute%d-air-but-more-countries-are-taking-action

[6] P. Kassomenos, M. Petrakis, D. Sarigiannis, A. Gotti, and S. Karakitsios, "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*, vol. 4, nos. 3–4, pp. 263–272, Dec. 2011.

[7] Z. Sun, X. An, Y. Tao, and Q. Hou, "Assessment of population exposure to PM_{10} for respiratory disease in Lanzhou (China) and its health-related economic costs based on GIS," *BMC Public Health*, vol. 13, no. 1, p. 891, Dec. 2013.

[8] Y. Li, Z. Ma, C. Zheng, and Y. Shang, "Ambient temperature enhanced acute cardiovascular-respiratory mortality effects of $PM_{2.5}$ in Beijing, China," *Int. J. Biometeorology*, vol. 59, no. 12, pp. 1761–1770, Dec. 2015.

[9] G. Polezer, Y. S. Tadano, H. V. Siqueira, A. F. Godoi, C. I. Yamamoto, P. A. de André, T. Pauliquevis, M. de Fatima Andrade, A. Oliveira, P. H. Saldiva, P. E. Taylor, and R. H. Godoi, "Assessing the impact of $PM_{2.5}$ on respiratory disease using artificial neural networks," *Environ. Pollut.*, vol. 235, pp. 394–403, Apr. 2018.
[10] L. G. Ardiles, Y. S. Tadano, S. Costa, V. Urbina, M. N. Capucim, I. da Silva, A. Braga, J. A. Martins, and L. D. Martins, "Negative binomial regression model for analysis of the relationship between hospitalization and air pollution," *Atmos. Pollut. Res.*, vol. 9, no. 2, pp. 333–341, Mar. 2018.
[11] J. T. Belotti, D. S. Castanho, L. N. Araujo, L. V. da Silva, T. A. Alves, Y. S. Tadano, S. L. Stevan, F. C. Corráa, and H. V. Siqueira, "Air pollution epidemiology: A simplified generalized linear model approach optimized by bio-inspired metaheuristics," *Environ. Res.*, vol. 191, Dec. 2020, Art. no. 110106.
[12] L. N. Araujo, J. T. Belotti, T. A. Alves, Y. D. S. Tadano, and H. Siqueira, "Ensemble method based on artificial neural networks to estimate air pollution health risks," *Environ. Model. Softw.*, vol. 123, Jan. 2020, Art. no. 104567.
[13] K.-Y. Wang and T.-T. Chau, "An association between air pollution and daily outpatient visits for respiratory disease in a heavy industry area," *PLoS ONE*, vol. 8, no. 10, Oct. 2013, Art. no. e75220.
[14] S. Genc, Z. Zadeoglulari, S. H. Fuss, and K. Genc, "The adverse effects of air pollution on the nervous system," *J. Toxicol.*, vol. 2012, pp. 1–23, 2012.
[15] J. W. Kim, S. Park, C. W. Lim, K. Lee, and B. Kim, "The role of air pollutants in initiating liver disease," *Toxicol. Res.*, vol. 30, no. 2, pp. 65–70, Jun. 2014.
[16] F. S. D. A. Filho, F. Madeiro, S. M. M. Fernandes, P. S. G. D. M. Neto, and T. A. E. Ferreira, "Time-series forecasting of pollutant concentration levels using particle swarm optimization and artificial neural networks," *Química Nova*, vol. 36, no. 6, pp. 783–789, 2013.
[17] P. S. G. de Mattos Neto, G. D. C. Cavalcanti, F. Madeiro, and T. A. E. Ferreira, "An approach to improve the performance of PM forecasters," *PLoS ONE*, vol. 10, no. 9, Sep. 2015, Art. no. e0138507.
[18] P. S. G. de Mattos Neto, F. Madeiro, T. A. E. Ferreira, and G. D. C. Cavalcanti, "Hybrid intelligent system for air quality forecasting using phase adjustment," *Eng. Appl. Artif. Intell.*, vol. 32, pp. 185–191, Jun. 2014.
[19] A. Vlachogianni, P. Kassomenos, A. Karppinen, S. Karakitsios, and J. Kukkonen, "Evaluation of a multiple regression model for the forecasting of the concentrations of NO_x and PM_{10} in athens and helsinki," *Sci. Total Environ.*, vol. 409, no. 8, pp. 1559–1571, Mar. 2011.
[20] H. Niska, M. Rantamäki, T. Hiltunen, A. Karppinen, J. Kukkonen, J. Ruuskanen, and M. Kolehmainen, "Evaluation of an integrated modelling system containing a multi-layer perceptron model and the numerical weather prediction model HIRLAM for the forecasting of urban airborne pollutant concentrations," *Atmos. Environ.*, vol. 39, no. 35, pp. 6524–6536, Nov. 2005.
[21] J. Kukkonen, "Extensive evaluation of neural network models for the prediction of NO_2 and PM_{10} concentrations, compared with a deterministic modelling system and measurements in central Helsinki," *Atmos. Environ.*, vol. 37, no. 32, pp. 4539–4550, Oct. 2003.
[22] F. Biancofiore, M. Busilacchio, M. Verdecchia, B. Tomassetti, E. Aruffo, S. Bianco, S. Di Tommaso, C. Colangeli, G. Rosatelli, and P. Di Carlo, "Recursive neural network model for analysis and forecast of PM_{10} and $PM_{2.5}$," *Atmos. Pollut. Res.*, vol. 8, no. 4, pp. 652–659, Jul. 2017.
[23] S. P. Neuman, "Maximum likelihood Bayesian averaging of uncertain model predictions," *Stochastic Environ. Res. Risk Assessment*, vol. 17, no. 5, pp. 291–305, Nov. 2003.
[24] A. T. Sergio, T. P. F. de Lima, and T. B. Ludermir, "Dynamic selection of forecast combiners," *Neurocomputing*, vol. 218, pp. 37–50, Dec. 2016.
[25] P. R. A. Firmino, P. S. G. de Mattos Neto, and T. A. E. Ferreira, "Correcting and combining time series forecasters," *Neural Netw.*, vol. 50, pp. 1–11, Feb. 2014.
[26] T. Lux and L. Morales-Arias, "Forecasting volatility under fractality, regime-switching, long memory and student-t innovations," *Comput. Statist. Data Anal.*, vol. 54, no. 11, pp. 2676–2692, Nov. 2010.
[27] S. Yin, L. Liu, and J. Hou, "A multivariate statistical combination forecasting method for product quality evaluation," *Inf. Sci.*, vols. 355–356, pp. 229–236, Aug. 2016.
[28] F. Ye, L. Zhang, D. Zhang, H. Fujita, and Z. Gong, "A novel forecasting method based on multi-order fuzzy time series and technical analysis," *Inf. Sci.*, vols. 367–368, pp. 41–57, Nov. 2016.
[29] R. T. A. de Oliveira, T. F. O. de Assis, P. R. A. Firmino, and T. A. E. Ferreira, "Copulas-based time series combined forecasters," *Inf. Sci.*, vol. 376, pp. 110–124, Jan. 2017.
[30] S. Panigrahi and H. Behera, "A hybrid ETS–ANN model for time series forecasting," *Eng. Appl. Artif. Intell.*, vol. 66, pp. 49–59, Nov. 2017.
[31] R. T. Clemen, "Combining forecasts: A review and annotated bibliography," *Int. J. Forecasting*, vol. 5, no. 4, pp. 559–583, Jan. 1989.
[32] D. I. Jeong and Y.-O. Kim, "Combining single-value streamflow forecasts—A review and guidelines for selecting techniques," *J. Hydrol.*, vol. 377, pp. 284–299, Oct. 2009.
[33] K. F. Wallis, "Combining forecasts—Forty years later," *Appl. Financial Econ.*, vol. 21, nos. 1–2, pp. 33–41, Jan. 2011.
[34] R. Dell'Aquila and E. Ronchetti, "Stock and bond return predictability: The discrimination power of model selection criteria," *Comput. Statist. Data Anal.*, vol. 50, no. 6, pp. 1478–1495, Mar. 2006.
[35] J. Belotti, H. Siqueira, L. Araujo, S. L. Stevan, P. S. G. de Mattos Neto, M. H. N. Marinho, J. F. L. de Oliveira, F. Usberti, M. D. A. Leone Filho, A. Converti, and L. A. Sarubbo, "Neural-based ensembles and unorganized machines to predict streamflow series from hydroelectric plants," *Energies*, vol. 13, no. 18, p. 4769, Sep. 2020.
[36] N. Kourentzes, D. K. Barrow, and S. F. Crone, "Neural network ensemble operators for time series forecasting," *Expert Syst. Appl.*, vol. 41, no. 9, pp. 4235–4244, Jul. 2014.
[37] K. Siwek and S. Osowski, "Improving the accuracy of prediction of PM_{10} pollution by the wavelet transformation and an ensemble of neural predictors," *Eng. Appl. Artif. Intell.*, vol. 25, no. 6, pp. 1246–1258, Sep. 2012.
[38] R. M. S. Souza, G. P. Coelho, A. E. A. da Silva, and S. A. Pozza, "Using ensembles of artificial neural networks to improve PM_{10}forecasts," *Chem. Eng. Trans.*, vol. 43, pp. 2161–2166, 2015.
[39] E. Debry and V. Mallet, "Ensemble forecasting with machine learning algorithms for ozone, nitrogen dioxide and PM_{10} on the Prev'Air platform," *Atmos. Environ.*, vol. 91, pp. 71–84, Jul. 2014.
[40] G. E. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, vol. 734. Hoboken, NJ, USA: Wiley, 2011.
[41] S. S. Haykin, *Adaptive Filter Theory*. London, U.K.: Pearson, 2008.
[42] S. S. Haykin, *Neural Networks and Learning Machines*. New York, NY, USA: Prentice-Hall, 2009.
[43] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, vol. 2. Jul. 2004, pp. 985–990.
[44] H. Jaeger, "Short term memory in echo state networks," GMD-German Nat. Res. Inst. Comput. Sci., Darmstadt, Germany, GMD Rep. 152, 2002.
[45] M. Sugeno and T. Yasukawa, "A fuzzy-logic-based approach to qualitative modeling," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 1, pp. 7–31, Feb. 1993.
[46] M. R. Najafi and H. Moradkhani, "Ensemble combination of seasonal streamflow forecasts," *J. Hydrol. Eng.*, vol. 21, no. 1, Jan. 2016, Art. no. 04015043.
[47] G. Polezer, A. Oliveira, S. Potgieter-Vermaak, A. F. L. Godoi, R. A. F. de Souza, C. I. Yamamoto, R. V. Andreoli, A. S. Medeiros, C. M. D. Machado, E. O. dos Santos, P. A. de André, T. Pauliquevis, P. H. N. Saldiva, S. T. Martin, and R. H. M. Godoi, "The influence that different urban development models has on $PM_{2.5}$ elemental and bioaccessible profiles," *Sci. Rep.*, vol. 9, no. 1, Dec. 2019, Art. no. 14846.
[48] H. Siqueira, I. Luna, T. A. Alves, and Y. de Souza Tadano, "The direct connection between Box & Jenkins methodology and adaptive filtering theory," *Math. Eng., Sci. Aerosp.*, vol. 10, no. 1, pp. 27–40, 2019.
[49] M. Khashei and M. Bijari, "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 2664–2675, Mar. 2011.
[50] A. B. Sanchez, C. Ordonez, F. S. Lasheras, F. J. de Cos Juez, and J. Roca-Pardinas, "Forecasting SO_2 pollution incidents by means of Elman artificial neural networks and ARIMA Model," *Abstract Appl. Anal.*, vol. 2013, no. 6, p. 238259, 2013.
[51] H. Siqueira, L. Boccato, R. Attux, and C. Lyra, "Unorganized machines for seasonal streamflow series forecasting," *Int. J. Neural Syst.*, vol. 24, no. 3, May 2014, Art. no. 1430009.
[52] P. Santos, M. Macedo, E. Figueiredo, C. J. Santana, F. Soares, H. Siqueira, A. Maciel, A. Gokhale, and C. J. A. Bastos-Filho, "Application of PSO-based clustering algorithms on educational databases," in *Proc. IEEE Latin Amer. Conf. Comput. Intell.*, Nov. 2017, pp. 1–6.
[53] E. D. P. Puchta, R. Lucas, F. R. V. Ferreira, H. V. Siqueira, and M. S. Kaster, "Gaussian adaptive PID control optimized via genetic algorithm applied to a step-down DC-DC converter," in *Proc. 12th IEEE Int. Conf. Ind. Appl. (INDUSCON)*, Nov. 2016, pp. 1–6.

[54] E. D. P. Puchta, H. V. Siqueira, and M. D. S. Kaster, "Optimization tools based on metaheuristics for performance enhancement in a Gaussian adaptive PID controller," *IEEE Trans. Cybern.*, vol. 50, no. 3, pp. 1185–1194, Mar. 2020.

[55] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proc. 6th Int. Symp. Micro Mach. Human Sci.*, 1995, pp. 39–43.

[56] A. V. Oppenheim, A. S. Willsky, and H. Nawab, *Signals and Systems*. London, U.K.: Pearson, 1996.

[57] H. Siqueira, L. Boccato, I. Luna, R. Attux, and C. Lyra, "Performance analysis of unorganized machines in streamflow forecasting of Brazilian plants," *Appl. Soft Comput.*, vol. 68, pp. 494–506, Jul. 2018.

[58] R. S. Ettouney, F. S. Mjalli, J. G. Zaki, M. A. El-Rifai, and H. M. Ettouney, "Forecasting of ozone pollution using artificial neural networks," *Manage. Environ. Qual., Int. J.*, vol. 20, no. 6, pp. 668–683, 2009.

[59] R. Adhikari and R. K. Agrawal, "A combination of artificial neural network and random walk models for financial time series forecasting," *Neural Comput. Appl.*, vol. 24, no. 6, pp. 1441–1449, May 2014.

[60] P. S. G. de Mattos Neto, M. H. N. Marinho, H. Siqueira, Y. de Souza Tadano, V. Machado, T. Antonini Alves, J. F. L. de Oliveira, and F. Madeiro, "A methodology to increase the accuracy of particulate matter predictors based on time decomposition," *Sustainability*, vol. 12, no. 18, p. 7310, Sep. 2020.

[61] Y. Kachba, D. M. D. G. Chiroli, J. T. Belotti, T. A. Alves, Y. de Souza Tadano, and H. Siqueira, "Artificial neural networks to estimate the influence of vehicular emission variables on morbidity and mortality in the largest metropolis in south america," *Sustainability*, vol. 12, no. 7, p. 2621, Mar. 2020.

[62] E. P. dos Santos and F. J. Von Zuben, "Improved second-order training algorithms for globally and partially recurrent neural networks," in *Proc. Int. Joint Conf. Neural Netw. Process.*, 1999, pp. 1501–1506.

[63] H. Siqueira and I. Luna, "Performance comparison of feedforward neural networks applied to streamflow series forecasting," *Math. Eng., Sci. Aerosp.*, vol. 10, no. 1, pp. 41–53, 2019.

[64] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, vol. 344. New York, NY, USA: Wiley, 2009.

[65] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.

[66] H. V. Siqueira, L. Boccato, R. R. F. Attux, and C. L. Filho, "Echo state networks and extreme learning machines: A comparative study on seasonal streamflow series prediction," in *Neural Information Processing* (Lecture Notes in Computer Science), vol. 7664, 2012, pp. 491–500.

[67] H. V. Siqueira, L. Boccato, R. R. F. Attux, and C. L. Filho, "Echo state networks for seasonal streamflow series forecasting," in *Intelligent Data Engineering and Automated Learning—IDEAL* (Lecture Notes in Computer Science), vol. 7435, 2012, pp. 226–236.

[68] J.-S. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 3, pp. 665–685, Jun. 1993.

[69] K. Polat and S. S. Durduran, "Usage of output-dependent data scaling in modeling and prediction of air pollution daily concentration values (PM_{10}) in the city of Konya," *Neural Comput. Appl.*, vol. 21, no. 8, pp. 2153–2162, Nov. 2012.

[70] M. Oprea, S. F. Mihalache, and M. Popescu, "A comparative study of computational intelligence techniques applied to $PM_{2.5}$ air pollution forecasting," in *Proc. 6th Int. Conf. Comput. Commun. Control (ICCCC)*, May 2016, pp. 103–108.

[71] J. Jang and N. Gulley, *Fuzzy Logic Toolbox 2: User's Guide MATLAB*. Natick, MA, USA: The Mathworks, 2007.

[72] K. Siwek, S. Osowski, and M. Sowinski, "Evolving the ensemble of predictors model for forecasting the daily average PM_{10}," *Int. J. Environ. Pollut.*, vol. 46, pp. 199–215, 2011.

[73] J. Mendes-Moreira, C. Soares, A. M. Jorge, and J. F. D. Sousa, "Ensemble approaches for regression: A survey," *ACM Comput. Surv.*, vol. 45, no. 1, pp. 1–40, Nov. 2012.

[74] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Jan. 2003.

[75] H. Siqueira, M. Macedo, Y. D. S. Tadano, T. A. Alves, S. L. Stevan, D. S. Oliveira, M. H. N. Marinho, P. S. G. D. M. Neto, J. F. L. D. Oliveira, I. Luna, M. D. A. L. Filho, L. A. Sarubbo, and A. Converti, "Selection of temporal lags for predicting riverflow series from hydroelectric plants using variable selection methods," *Energies*, vol. 13, no. 16, p. 4236, Aug. 2020.

[76] R. B. Nelsen, *An Introduction to Copulas*, 2nd ed. Portland, ON, USA: Springer, 2006.

[77] A. Sklar, "Fonctions de repartition a n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, vol. 8, pp. 229–231, Dec. 1959.

[78] A. Patton, "Copula methods for forecasting multivariate time series," in *Handbook Economic Forecasting*, vol. 2. Amsterdam, The Netherlands: Elsevier, 2013, pp. 899–960.

[79] A. J. Patton, "A review of copula models for economic time series," *J. Multivariate Anal.*, vol. 110, pp. 4–18, Sep. 2012.

[80] R. T. A. Oliveira, T. F. O. Assis, P. R. A. Firmino, T. A. E. Ferreira, and A. L. I. de Oliveira, "Copulas-based ensemble of artificial neural networks for forecasting real world time series," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2016, pp. 4089–4096.

[81] R. T. A. D. Oliveira, T. F. Oliveira, P. R. A. Firmino, and T. A. E. Ferreira, "Combining time series forecasting models via Gumbel-Hougaard copulas," in *Proc. 11th Brazilian Congr. Comput. Intell.*, Sep. 2013, pp. 1–6.

[82] T. F. Oliveira, P. R. Firmino, and T. A. E. Ferreira, "Study of models' uncertainty by Gumbel copula (in Portuguese)," in *Proc. 58th RBras Reunião Ann. Região Brasileira Soc. Int. Biometria*, 2013, pp. 1–13.

[83] H. Joe and D. Kurowicka, *Dependence Modeling: Vine Copula Handbook*. Singapore: World Scientific, 2011.

[84] T. F. Oliveira, R. T. A. D. Oliveira, P. R. A. Firmino, P. S. G. D. M. Neto, and T. A. E. Ferreira, "Combination of biased artificial neural network forecasters," in *Proc. 11th Brazilian Congr. Comput. Intell.*, Sep. 2013, pp. 1–6.

[85] Statistics in Finland. *Population Projection 2019: Vital Statistics by Sex and Area 2019–2040*. Accessed: Jan. 25, 2020. [Online]. Available: https://pt.knoema.com/statfin_vaenn_pxt_128w/population-projection-2019-vital-statistics-by-sex-and-area-2019-2040-finland

[86] W. Spark. (Aug. 2020). *Mean Meteorological Conditions of Campinas, São Paulo, Ipojuca, Helsinki, and Region (in Portuguese: Condições Meteorológicas Médias de Campinas, São Paulo, Ipojuca, Helsinki e Região)*. Accessed: Aug. 22, 2020. [Online]. Available: http://pt.weatherspark.com

[87] D. Voukantsis, K. Karatzas, J. Kukkonen, T. Räsänen, A. Karppinen, and M. Kolehmainen, "Intercomparison of air quality data using principal component analysis, and forecasting of PM_{10} and $PM_{2.5}$ concentrations using artificial neural networks, in Thessaloniki and Helsinki," *Sci. Total Environ.*, vol. 409, no. 7, pp. 1266–1276, Mar. 2011.

[88] IBGE-Brazilian Institute of Geography and Statistics (in Portuguese: Instituto Brasileiro de Geografia Estatística. *Censo 2010*. Accessed: Aug. 22, 2019. [Online]. Available: https://censo2010.ibge.gov.br/

[89] L. Prechelt, "Proben1: A set of neural network benchmark problems and benchmarking rules," Univ. Karlsruhe, Karlsruhe, Germany, Tech. Rep. 21/94, 1994.

[90] A. R. Lima Junior, D. A. Silva, P. S. Mattos Neto, and T. A. E. Ferreira, "An experimental study of fitness function and time series forecasting using artificial neural networks," in *Proc. 12th Annu. Conf. comp Genetic Evol. Comput. (GECCO)*, 2010, pp. 2015–2018.

[91] P. S. G. de Mattos Neto, A. R. Lima, T. A. E. Ferreira, and G. D. C. Cavalcanti, "An intelligent perturbative approach for the time series forecasting problem," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2010, pp. 1–8.

[92] G. W. Corder and D. I. Foreman, *Nonparametric Statistics for Non-Statisticians*. Hoboken, NJ, USA: Wiley, 2009.

[93] L. N. De Castro, *Fundamentals Natural Computing: Basic Concepts, Algorithms, Application*. Boca Raton, FL, USA: CRC Press, 2006.

[94] Y. de Souza Tadano, H. V. Siqueira, and T. A. Alves, "Unorganized machines to predict hospital admissions for respiratory diseases," in *Proc. IEEE Latin Amer. Conf. Comput. Intell. (LA-CCI)*, Nov. 2016, pp. 1–6.

[95] R. C. Team, *A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2020.

[96] D. Dunea, A. Pohoata, and S. Iordache, "Using wavelet–feedforward neural networks to improve air pollution forecasting in urban environments," *Environ. Monitor. Assessment*, vol. 187, no. 7, p. 477, Jul. 2015.

[97] Y. D. S. Tadano, H. V. Siqueira, T. A. Alves, and M. H. D. N. Marinho, "Forecasting particulate matter concentrations: Use of unorganized machines," *Int. J. Adv. Eng. Res. Sci.*, vol. 4, no. 4, pp. 188–191, 2017.

[98] X. Feng, Q. Li, Y. Zhu, J. Hou, L. Jin, and J. Wang, "Artificial neural networks forecasting of $PM_{2.5}$ pollution using air mass trajectory based geographic model and wavelet transformation," *Atmos. Environ.*, vol. 107, pp. 118–128, Apr. 2015.

[99] M. Madhiarasan and S. N. Deepa, "A novel criterion to select hidden neuron numbers in improved back propagation networks for wind speed forecasting," *Int. J. Speech Technol.*, vol. 44, no. 4, pp. 878–893, Jun. 2016.

[100] United Nations. (Aug. 2019). *Sustainable Development Goals: Knowledge Platform*. Accessed: Aug. 25, 2020. [Online]. Available: https://sustainabledevelopment.un.org/

PAULO S. G. DE MATTOS NETO received the Ph.D. degree in computer science from the Center of Informatics, Federal University of Pernambuco, in 2012. He is currently a Researcher with the Brazilian National Research Council. He is also a Professor with the Center of Informatics of the Federal University of Pernambuco. His current research interests include time series modeling, artificial neural networks, and hybrid systems. He is a member of the International Neural Network Society and the Brazilian Computer Science Society.

PAULO RENATO A. FIRMINO received the B.S. degree in statistics, and the M.S. and Ph.D. degrees in production engineering (operational research) from the Federal University of Pernambuco, in 2002, 2004, and 2009, respectively. He is currently a Professor with the Center for Science and Technology, Federal University of Cariri, Brazil. His research interest includes quantitative risk assessment, with a focus on modeling, simulation, and forecasting.

HUGO SIQUEIRA received the B.Sc. degree in electrical engineering from the State University of São Paulo, in 2006, and the master's and Ph.D. degrees from the University of Campinas, in 2009 and 2013, respectively. He was a Post-Doctoral Researcher with the University of Campinas, in 2014, and Illinois State University, USA, and the University of Pernambuco, in 2017. He has worked on electric engineering and computer science, with an emphasis on neural networks, evolutionary algorithms, immune algorithms, swarm intelligence, time series forecasting, pollutant impact on human health, and clustering tasks, among others. He is currently an Adjunct Professor with the Federal University of Technology—Paraná (UTFPR). He is also an Adviser of Graduate Programs in computer science (PPGCC) and a Production Engineering (PPGEP). He is the Coordinator of the Interdisciplinary Group of Computational Intelligence and the Laboratory of Computational Intelligence and Advanced Control, UTFPR-PG.

YARA DE SOUZA TADANO received the B.Sc. degree in physics from the Federal University of Mato Grosso do Sul, in 2004, the master's degree from the Federal University of Technology—Paraná (UTFPR), in 2007, and the Ph.D. degree in mechanical engineering from the University of Campinas, in 2012. She was a Postdoctoral Researcher in environmental engineering with UTFPR, in 2013. She is currently a Professor with UTFPR and also an Adviser of the Postgraduate Program (master's degree) in mechanical engineering. She is the Coordinator of the Research network in Life Cycle Impact Assessment and the Laboratory of Air Pollutants Dispersion. Her researches are related to air pollution, such as emission and dispersion of pollutants, environmental impact, population health, statistical analysis, artificial neural networks, and life cycle impact assessment.

THIAGO ANTONINI ALVES received the D.Sc. degree in mechanical engineering from the State University of Campinas, in 2010. He is currently a Professor of mechanical engineering with the Federal University of Technology—Paraná (UTFPR), Ponta Grossa, Brazil. He has experience in the field of thermal engineering, focusing on heat and mass transfer, thermodynamics, and energy, with an emphasis on heat pipes and thermosyphons, convection, conduction, electronic cooling, solar energy, generation and cogeneration of energy, emission and dispersion of air pollutants, numerical simulations, and experimental analysis.

JOÃO FAUSTO L. DE OLIVEIRA received the Ph.D. degree in computational intelligence from the Federal University of Pernambuco, in 2016. He joined the University of Pernambuco, in 2014, where he is currently a Professor and the Coordinator of the Automation and Control Engineering undergraduate program. His research interests include artificial neural networks, pattern recognition, hybrid intelligent systems, evolutionary computation, and time series forecasting.

MANOEL HENRIQUE DA NÓBREGA MARINHO received the bachelor's degree in civil engineering from the Federal University of Campina Grande, in 1999, and the master's degree in civil engineering and the Ph.D. degree in electrical engineering from UNICAMP, in 2002 and 2005, respectively. He has coordinated and participated in several research and development projects in renowned companies in the area of electricity generation and distribution. He is currently an Adjunct Professor with the Polytechnic School of Engineering, University of Pernambuco, and a Permanent Professor with the Systems Engineering Graduate Program, teaching disciplines, supervising graduate and undergraduate students, in addition to teaching undergraduate classes in engineering and materials physics. He is also in related with in research and development projects and involved in teaching and extension activities in the areas, such as probability, statistics, stochastic processes, water resources, time series forecasting, computational intelligence, smart meters, renewable energy, energy storage, planning and programming of the operation of electric power systems, hydrothermal power systems, energy planning, reliability engineering in life data analysis, and repairable systems and extreme conditions of use.

FRANCISCO MADEIRO received the D.Sc. degree in electrical engineering from the Federal University of Paraiba, Brazil, in 2001. Since 2006, he has been with the University of Pernambuco, Brazil, where he is currently an Associate Professor. His main research interests include signal processing, communications systems, and computational intelligence.