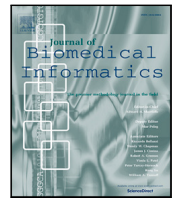Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Original Research

Multi-step ahead meningitis case forecasting based on decomposition and multi-objective optimization methods

Matheus Henrique Dal Molin Ribeiro [a,b,*], Viviana Cocco Mariani [c,d], Leandro dos Santos Coelho [a,c]

[a] *Graduate Program in Industrial & Systems Engineering (PPGEPS), Pontifical Catholic University of Parana (PUCPR), 1155, Rua Imaculada Conceicao, Curitiba, Parana, 80215-901, Brazil*

[b] *Department of Mathematics, Federal Technological University of Parana (UTFPR), Via do Conhecimento, KM 01 - Fraron, Pato Branco, Parana, 85503-390, Brazil*

[c] *Department of Electrical Engineering, Federal University of Parana (UFPR), 100, Avenida Cel. Francisco dos Santos, Curitiba, Parana, 81530-000, Brazil*

[d] *Mechanical Engineering Graduate Program (PPGEM), Pontifical Catholic University of Parana (PUCPR), 1155, Rua Imaculada Conceicao, Curitiba, Parana, 80215-901, Brazil*

ARTICLE INFO

ABSTRACT

Epidemiological time series forecasting plays an important role in health public systems, due to its ability to allow managers to develop strategic planning to avoid possible epidemics. In this paper, a hybrid learning framework is developed to forecast multi-step-ahead (one, two, and three-month-ahead) meningitis cases in four states of Brazil. First, the proposed approach applies an ensemble empirical mode decomposition (EEMD) to decompose the data into intrinsic mode functions and residual components. Then, each component is used as the input of five different forecasting models, and, from there, forecasted results are obtained. Finally, all combinations of models and components are developed, and for each case, the forecasted results are weighted integrated (WI) to formulate a heterogeneous ensemble forecaster for the monthly meningitis cases. In the final stage, a multi-objective optimization (MOO) using the Non-Dominated Sorting Genetic Algorithm – version II is employed to find a set of candidates' weights, and then the Technique for Order of Preference by similarity to Ideal Solution (TOPSIS) is applied to choose the adequate set of weights. Next, the most adequate model is the one with the best generalization capacity out-of-sample in terms of performance criteria including mean absolute error (MAE), relative root mean squared error (RRMSE), and symmetric mean absolute percentage error (sMAPE). By using MOO, the intention is to enhance the performance of the forecasting models by improving simultaneously their accuracy and stability measures. To access the model's performance, comparisons based on metrics are conducted with: (i) EEMD, heterogeneous ensemble integrated by direct strategy, or simple sum; (ii) EEMD, homogeneous ensemble of components WI; (iii) models without signal decomposition. At this stage, MAE, RRMSE, and sMAPE criteria as well as Diebold–Mariano statistical test are adopted. In all twelve scenarios, the proposed framework was able to perform more accurate and stable forecasts, which showed, on 89.17% of the cases, that the errors of the proposed approach are statistically lower than other approaches. These results showed that combining EEMD, heterogeneous ensemble and WI with weights obtained by optimization can develop precise and stable forecasts. The modeling developed in this paper is promising and can be used by managers to support decision making.

1. Introduction

Meningitis is an inflammation that has several classifications with specific causes and symptoms, and, unfortunately, it is still a major public health problem because it causes irreversible health damages and has kept mortality rates high [1]. Besides, this disease could be a result of viral, bacterial, or fungal infection of the fluid surrounding the brain and spinal cord [2]. Early diagnosis and the immediate initiation of the treatment are fundamental for a good prognosis of the disease. Concerning the number of meningitis cases registered in Brazilian states located in the South or Southeast such as Parana (PR), São Paulo (SP), Minas Gerais (MG), and Rio de Janeiro (RJ) presented over ten thousand cases in 2018 [3]. Within this context, the development of

---

efficient public policies and preventive campaigns are the key to change this scenario. The development of efficient predictive models plays an important role in epidemiological modeling.

Predictive mathematics models are widely used to design the epidemiological scenario of specific diseases. These approaches are adopted for different purposes such as comparing, implementing, evaluating, prevention, therapy, and the development of public policies [4]. Usually, the epidemic models are based on parameters related to susceptibility (S), infected (I), and removed (R), as well as exposed (E) individuals can be considered which leads to the SIR or SEIR models. Each variation of these models has its particularity, and different factors can be considered in these approaches to provide knowledge about the disease spread. Nowadays, these approaches have been proposed to understand the spread of new coronavirus [5,6]. Also, different mathematical approaches are proposed to mitigate the effects of several diseases such as ebola [7], influenza [8], and malaria [9]. In the last years, a computer science field called artificial intelligence (AI) able to recognize patterns of historical data and support the decision making has received attention to solving problems from the commerce [10], and industry [11]. Machine learning models, an AI sub-field, becomes the kernel of data analysis, once dealing with classification [12], data clustering [13], and regression tasks [14]. Nonetheless, when it comes to matters of diseases that plague the Brazilian public health system, such as dengue, malaria, and others, there are limited discussions as regards the effectiveness of machine learning models to develop predictive models. Some of these studies aimed to define the incidence of diseases such as *ventriculitis* and meningitis [15], as well as map the transmission risk of Zika [16], and make probabilistic forecasting of influenza [17].

Considering the purpose of knowing the number of future cases for any disease, once the datasets contain temporal information, time series forecasting should be used for this task. Therefore, time series forecasting aims to use past data to know future values with the purpose of, for example, making strategic planning to improve the knowledge in the domains that are inserted and help develop public policies. In this aspect, developing an efficient model is desirable; moreover, techniques such as the ensemble learning and decomposition may be used for this purpose. These strategies can be employed to deal with nonlinearity, nonstationarity, and cyclicity inherent to time series.

Faced with this, the ensemble learning is an approach that is applicable in regression [18] and/or classification [19] tasks which have the objective of improving the model's predictive accuracy. The main aspect of this approach lies in training several base (weak) models and combining their predictions to build an efficient model [20]. It is believed that this improvement occurs because each base model learns the different characteristics of the data and adds this information to the final results. Indeed, this methodology has proven effective in forecasting tasks in different fields of knowledge such as agribusiness [21], chemometrics [22], ecology [23], and medicine [24,25].

Also, an additional approach usually adopted to improve the models' performance is the ensemble empirical mode decomposition (EEMD) [26]. By employing EEMD, it is possible to separate the original signal into components (intrinsic mode functions — IMF and residual signals) with different amplitudes and frequencies added to the noise, aiming to extract relevant data information. The EEMD approach has been proven effective in then forecasting field of dealing with applications in several domains of knowledge, such as aquaculture [27], economy [28], and energy [29]. The EEMD steps consist of data decomposition and aggregating the components or aggregating the predictions of each component to obtain the original signal. In respect of this, each component is treated as an input set, and they can be trained separately using various algorithms. One strategy considers the same weight for all components (direct strategy — DI), which penalizes those that explain more data variability by attributing the same importance for all components, while another strategy employs different weights (weighted strategy – WI) for each component. Because

the disease series show up nonlinear and nonstationary on most of the cases, the EEMD is efficient to analyze series related to meningitis cases.

Considering the aforementioned, because of the use of the EEMD approach, two questions emerge. The first one lies in which algorithm should be employed to train and forecast each component. The second refers to which approach will be used to reconstruct the decomposed signal, that is, which strategy should be adopted to obtain an ensemble of components. To deal with these questions, this paper aims to propose a hybrid framework that was developed in three steps for multi-month-ahead forecasting (one, two and three-month-ahead) meningitis cases in the Brazilian states of MG, SP, PR, and RJ. As regards to this, in the first, the original signal is split into five components (four IMF and one residual). In the sequence, each component is used as the input of techniques: Bayesian Regularized Neural Networks (BRNN) [30], Cubist (CUBIST) [31], Gradient Boosting Machine (GBM) [32], Partial Least Squares (PLS) [33], and Quantile Random Forest (QRF) [34]. Following, predictions for all components are obtained by each model. Different combinations of models to predict the components are developed, and, for each combination, the results are weighted integrated to formulate a heterogeneous ensemble (HTE), considering that at least two different models are used in each combination. Besides, multi-objective optimization (MOO) using elitist Non-Dominated Sorting Genetic Algorithm — version II (NSGA-II) [35] is employed to find a Pareto Front (composed by a set of candidate's weight), and Technique for Order of Preference by similarity to Ideal Solution (TOPSIS) [36] is applied to choose the adequate set of weights from Pareto Front. Finally, considering all the combinations developed, the most adequate model is the one with the best generalization of out-of-sample capacity in terms of mean absolute error (MAE), relative root mean squared error (RRMSE), and symmetric mean absolute percentage error (sMAPE).

The contributions of this paper lie in four-folds, such as:

1. The development of a unique hybrid framework based on data decomposing, ensemble, and MOO can improve simultaneously the accuracy and stability of forecasts of meningitis cases.
2. Performing multi-step-ahead forecast, once that in papers related to diseases incidence forecasting, single forecasting horizon is used;
3. An investigation of the effectiveness of the multi-objective procedure for choosing weights used to aggregate the forecasts of EEMD components. In this regard, this paper seeks to add discussions in this field, taking into account most of the debates regarding single–objective optimization for components aggregation, such as Du et al. [37] and Zhang and Wang [38];
4. Another contribution lies in the use of a diversified set of algorithms, concerning algorithms structure, to train and forecast each component after the EEMD performance, where these different algorithms can capture their inherent variability.

The remainder of this paper is structured as follows: Section 2 presents some related works on the use of machine learning to forecasting disease cases. Section 3 describes the datasets adopted, while Section 4 details the methods employed in this paper for such an approach. Section 5 lists the steps for modeling. Section 6 presents the results and discussions. Finally, Section 7 concludes and presents the proposals of future research for the theme hereby adopted.

2. Related works

This section presents a summary of some recent developments in time series forecasting for epidemic diseases using machine learning and general models. Considering the matter above mentioned, there were sixteen relevant papers, in the last years, found in the literature. With regard to diseases type, most of the papers aimed to perform comparisons either between machine learning models on the task of
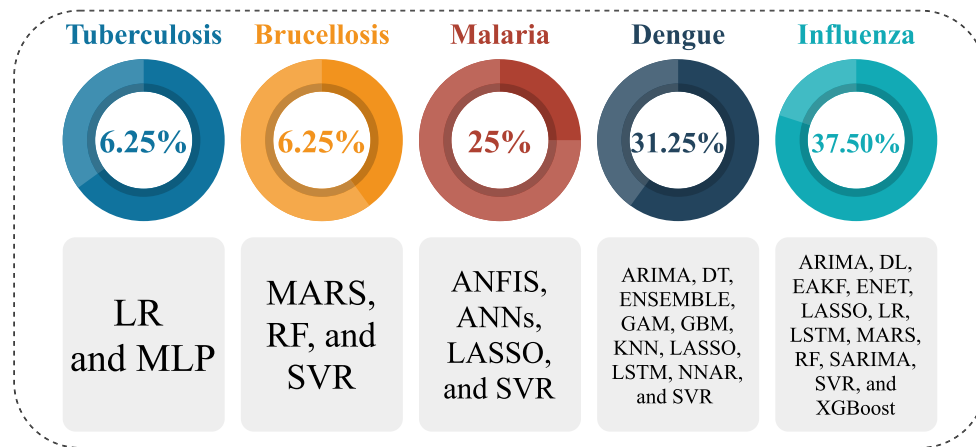
Fig. 1. Diseases related and its adopted modeling.

forecasting future cases or pandemic risk for dengue, malaria, and influenza. In these cases, one paper focuses on the diseases of dengue and malaria, simultaneously. This information is summarized in Table A.1 contained in Appendix A which presents a review of related papers with the use of machine learning and general models for forecasting disease cases.

In the context of the employed techniques, the adopted approaches may be split into different classes, such as: (i) Artificial neural networks (ANNs): Multilayer perceptron (MLP), deep learning (DL), neural network autoregressive (NNAR) model, and long-short-term memory (LSTM); (ii) Regression trees: Decision trees (DT), random forests (RF), gradient boosting machine (GBM), and eXtreme gradient boosting (XGBoost); (iii) General regressions: Least absolute shrinkage and selection operator (LASSO), support vector regression (SVR), multivariate adaptive regression splines (MARS), elasticnet (ENET), and linear regression (LR); (iv) Remain approaches: Ensembled adjustment Kalman filter (EAKF), Gaussian process (GP), generalized additive model (GAM), autoregressive integrated moving average (ARIMA), and seasonal autoregressive integrated moving average (SARIMA). Fig. 1 associates the diseases and adopted modeling.

Adjacent to the above mentioned, as well the presented in Appendix A, some gaps in relation to the developed approaches can be found and are stated as follows:

- Considering the disease's type, around of 93.75% of the papers focused on malaria, dengue or influenza. Hence, there is a lack of discussion concerning the predictive capacity of machine learning-based approaches for diseases such as measles, meningitis, and chikungunya on the forecast task;
- In the modeling aspect, only four papers focused on ensemble approaches such as bagging and boosting or models combined by average. One paper used an optimization approach of the swarm intelligence field called firefly algorithm (FFA) for hyperparameters tuning, and no paper adopted signal decomposition or MOO with the purpose of building ensembles. It is well known that the combination of these strategies can help on the improvement of the model's accuracy and, therefore, out-of-sample generalization;
- In general, to the forecast horizon, most of the papers focused on a single horizon, that is, one week or one-month-ahead. The discussions are narrow for the multi-step-ahead forecast. Indeed, the use of other horizons could be important because they can be useful for managers to develop strategic plains with the goal of preventing the population from having these diseases, as well as developing strategies and public policies so that these diseases will affect as few people as possible.
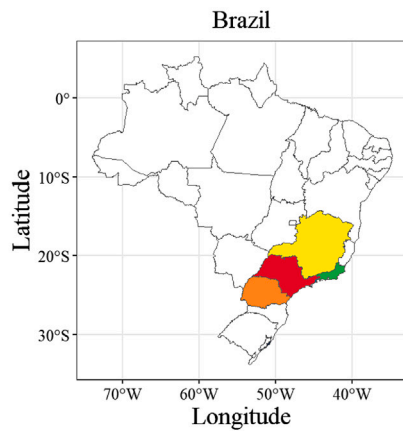
Therefore, faced with the aforementioned aspects, this study seeks to fill the pointed-out gaps, by proposing a hybrid framework to forecast the number of meningitis cases, which considers EEMD, heterogeneous ensemble, and MOO approaches to reconstruct the decomposed signal.

3. Material

The datasets adopted in this paper refer to the monthly number of confirmed cases of meningitis, recorded on a disease information system at the Brazilian Ministry of Health. The adopted period of analysis ranges from January of 2007 to December of 2018, and the data is available at the Department of Informatics of the Unified Health System (*Departamento de Informática do Sistema Único de Saúde*, in Portuguese, DATASUS) [3]. The data ranged from January of 2007 to December of 2017 is used in the training process and the remaining information is used to test the proposed framework's performance, respectively. In this context, the adopted data refers to information on the number of confirmed cases of meningitis in the Brazilian states PR, SP, MG, and RJ taking into account that these states presented the four largest notifications number in 2018.

In this paper, to verify the generalization ability of the proposed methodology, the above-described datasets are used. We did not fit the proposed model on the time series of a single region, and then validate the model by evaluating the prediction of new cases of another region because each state has different features, such as demography, geography, population density, economic and human development index, as well as public health policies. Therefore, to avoid erroneous conclusions, the proposed model is trained (training set) and validated (test set, out-of-sample forecasting) in the same Brazilian state, but in different splitting setups of the datasets. In this way, we tried to accommodate the data variability for each state.

Fig. 2 shows the study areas, the behavior of the number of notified cases by state and its descriptive measures, as well as the autocorrelation function (ACF) for the MG(A), PR(B), RJ(C) and SP(D) series. Secondly, as highlighted by Fig. 2, for the state of PR, there is a greater number of notifications than presented for other states, as reported by the descriptive measures (number of observations — n, minimum — Min, average, maximum — Max, and standard deviation — SD). The Augmented Dickey–Fuller (DF) test shows that the four time series are non-stationary (DF = −5.35 - −3.32, *p*-value > 0.05). Aiming at evaluating the presence of seasonality within the data, the Kruskal–Wallis test is performed. In this case, for the MG, SP, and RJ time series, there is no evidence of seasonality (χ^2_{11} = 13.32 - 15.50, *p*-value > 0.05), while the series related to the state of PR present evidence of seasonality (χ^2_{11} = 33.07, *p*-value < 0.05) [39]. Additionally, the autocorrelation measures suggest, due to lags, that up to the first four observations are correlated and can be used as inputs for the data modeling.

Fig. 2. Study area representation and datasets behavior.

4. Methods

The objective of this section is to present each technique employed in this paper.

4.1. Ensemble Empirical Mode Decomposition

The EEMD technique was proposed by Wu and Huang [26]. It is an extension of the Empirical Mode Decomposition (EMD) algorithm and should be used with the purpose of data de-noising. The EMD approach allows the analyzing of nonlinear and non-stationary signals. The steps followed in this technique are data decomposition, and later data aggregation. First, in the decomposition stage, the EMD separates the original signal into coexisting oscillatory components (IMF and residual signal) [40]. Next, after the modeling of each component, it is necessary to perform the aggregation of results to obtain the original signal.

When the EMD is used, the main drawback is named "mode mixing", that is, each single IMF consists of signals with dramatically disparate scales or a signal of the same scale appears in different IMF

components [26]. Seeking to solve this problem, the noise assisted EEMD was proposed. In this approach, EMD is performed k times, and different white noise (a random signal that follows a normal distribution with zero mean and constant variance) is added to the data in each trial. Two disadvantages can be stated for EEMD, as follows: (i) extra noise exists in the reconstructed signal and (ii) it needs more computational resources than EMD [41].

Initially, the EMD approach is presented and generalized for EEMD. In fact, the main steps are stated as follows:

1. Input the original time series;
2. Add white noise on the original time series;
3. Obtain the local maximum and minimum values of time series of step 2;
4. Generate envelops for the new time series, taking into account its boundaries;
5. Calculate the average of the boundaries of step 4;
6. Extract from the generated data the average value obtained in step 5;
7. After performing each step 2–6, each component is named IMF [42] if it has the following characteristics: (i) in the entire dataset, the number of extreme and zero crossings must either be equal or different, at most, by one, and (ii) at any point, the mean value of the envelopes is defined by the local maximal and the local minimal which must be zero;
8. Defining that the component of the step 7 is an IMF, the residual component is computed as the difference between the IMF and the remaining data. The steps 2–7 are performed until that residual component becomes a monotonic function or has one local extreme point from which no more IMF can be extracted.

The EEMD performed the steps 2–8 k times and for each trial there are n components (IMF$_{(1,k)}$, ..., IMF$_{(n,k)}$, Residual$_{(n+1,k)}$). Therefore, each final component is treated as the average of the respective components of the k executions; the number of repeated procedures is called the ensemble number. After the end of the decomposition process, the original time series can be expressed as the sum of IMF and residual components. In this stage, DI or WI strategies should be used. The DI consists of assuming the same weight for each component, while WI considers different weights for each component. When a decomposition technique is adopted, there are two possibilities for data analysis, as follows: (i) datasets are decomposed and components are aggregated and used as input in some forecasting method, see the example developed by Jiang and Liu [43], and (ii) each component obtained during the decomposition stage is used as an input by some forecast method, and the forecasting of each component is used in the aggregation stage, see the example developed by Wang et al. [44] or Wu et al. [45]. In this paper, the second possibility and WI are used for data modeling. The parameters used for the decomposition task are presented in Section 5, item 1.

4.2. Bayesian Regularized Neural Network

The BRNN is a type of feedforward neural network, composed by one input and one hidden layer, which uses the Bayesian methods, such as Empirical Bayes, for parameter estimation, with the purpose of avoiding overfitting [46]. The mathematical modeling can be stated as follows:

$$y_i = \sum_{k=1}^{s} w_k g_k \left(b_k + \sum_{j=1}^{p} x_{ij} \beta_j^k \right) \tag{1}$$

in which y_i is ith output value ($i = 1, \ldots, n$), s is the number of neurons, x_{ij} is an input value, $\theta = [w_1, \ldots, w_k, b_1, \ldots, b_k, \beta_1^1, \ldots, \beta_p^1, \beta_1^s, \ldots, \beta_p^s]$, is the vector of weights and biases, w_k is the weight of kth neuron ($k = 1, \ldots, s$), b_k is the bias of kth neuron, β_j^k is the weight of jth input

to the network ($j = 1, \ldots, p$), and $g(.)_k$ is the activation function. In this paper, the activation function is the hyperbolic tangent.

According to the procedure proposed by MacKay [46], the elements of θ are estimated by using an empirical Bayesian approach, according to two steps described as follows,

1. Minimization of

$$F(\theta) = \frac{1}{2\sigma_e^2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \frac{1}{2\sigma_\theta^2} \sum_{j=1}^{m} \theta_j^2 \tag{2}$$

in which, $F(.)$ is a function of θ, y_i and $\hat{y}_i$ are the observed and estimated values, respectively. Also, σ_e^2 and σ_θ^2 are the variance of errors, weights and biases, respectively;
2. Updating the variance components σ_e^2 and σ_θ^2 by maximizing a conditional probability (p) given the observed values according to the variances, i.e., $p(\mathbf{y}|\sigma_e^2, \sigma_\theta^2)$. When the early expression is not present in the closed form, an approximation for the marginal log-likelihood is adopted.

In Eq. 2, the variances are regularization parameters, where the trade-off between goodness-of-fit and smoothing can be controlled. In this approach, only the neurons number should be determined.

4.3. CUBIST

The CUBIST is a rule-based model, which performs predictions following the regression of trees principle [47]. Giving a regression tree, for each leaf is constructed a linear regression model, or in this case, a rule, associated with the information contained in each leaf. Once all rules are constructed, the final predictions are based on the linear combination of developed rules. The main difference between a simple regression tree and a CUBIST model is how the models make predictions within the nodes [31]. In order, to attempt an improvement in the model's accuracy, this rule-based model generates a set of rules, named a committee, as happens with the boosting approach. This is developed to correct the predictions of the previous members. Also, given a set of new features, the CUBIST can adjust the model prediction using samples from the training set by employing the neighborhood concept, like the k-nearest-neighbors approach [48]. Aiming at finding the number of neighbors, a distance measure, in this case, Manhattan is applied. Finally, once the set of committees and number of neighbors are defined, the final rule prediction is the simple average of committees predictions.

4.4. Generalized boosted regression

The generalized boosted regression modeling or gradient boosting machine (GBM) model is based on the boosting principle, proposed by Friedman [49], which seeks to find an additive model that minimizes the loss function. By using a gradient descent approach, the GBM builds models in the negative sense of the partial derivative of the loss function regarding the set of predictors. The GBM performs iteratively so that a regression tree model is fitted to the data, and the residuals from the fitted model are obtained. A new model is adjusted for previous residuals, a new prediction is obtained, to which the initial forecast is added, and then a new residual measure is obtained. This process is performed iteratively until a convergence criterion is reached, and the final prediction is obtained by the average in regression problems. Adjacent to the improvement that can be reached by the GBM structure, this process tends to generate overfitting. To avoid this problem, weights are assigned for observations whose errors are greater. In this paper, for the adopted GBM approach, control hyperparameters such as number of boosting iterations, maximum tree depth, shrinkage and minimum terminal node size are used. The best combination of these parameters should be found; this is a challenge since an inappropriate choice can result in a computationally costly and inefficient process.

4.5. Quantile random forests

The quantile random forests (QRF) [50] approach is an extension of the RF ensemble learning model [51]. It provides information about the full conditional distribution of the response variable, and not only about the conditional mean. In this approach, the use of conditional quantile is to enhance the RF performance, which makes this a consistent approach [50]. The main assumption about QRF lies on the weighted observations that can be used for estimating the conditional average [52]. Additionally, while the RF approach retains the result's information about the average of notifications number of the leaves, the QRF keeps all notifications contained in the leaves to estimate, where

$$\hat{P}\left(y_{t+h} \leq y | p_t\right) = \hat{F}(y|p_t),\tag{3}$$

in which the left side of the equation represents the conditional probability of the number of notifications conditioned to the predictors, being y_{t+h} the notifications in a h horizon and p_t the predictors' number, and $(t = 1, \ldots, N)$. The right side refers to the conditional distribution function (CDF) of the average notification numbers regarding the predictors. Considering that QRF uses the quantiles in the prediction process, the α-quantile of CDF is stated as the probability that the number of notifications is lower than Q_α if the given p_t is equal to α, where the estimate of α is stated as follows:

$$\hat{Q}_\alpha(p_t) = \inf \left\{y : \hat{F}(y|p_t) \geq \alpha\right\},\tag{4}$$

in which $\hat{Q}_\alpha(p_t)$ is the α-quantile estimated relation to the predictor t [50]. For the QRF architecture adopted in this paper, only the number of randomly selected predictors should be determined.

4.6. Partial least squares

The PLS regression approach is a technique to analyze multivariate data, in which the aim is to relate one or two output variables (**Y**) with several inputs (**X**). For this purpose, given a linear model, the problem that often arises is the matrix of inputs being singular. Faced with this, to deal with this problem, the PLS decomposes **X** into orthogonal scores **T** and loadings **P**, in which $\mathbf{X} = \mathbf{T} \times \mathbf{P}$. Regarding this approach, **T** and **P** values are chosen in such a way that the covariance between inputs and outputs is maximized. In other words, the PLS finds components that maximize the variation of the predictors while simultaneously requiring these components to have a maximum correlation with the response. In a more general way, the PLS finds linear combinations of the predictors, named components that are like the principal component analysis. Additionally, the PLS in its classical form is based on the nonlinear iterative partial least squares [53,54], and the number of components should be defined.

4.7. Multi-objective optimization

On some problems, it is necessary to minimize (or maximize) multiple objectives to achieve a preferable solution. Naturally, MOO should be adopted [55], and it is performed in three steps. First, it is necessary to define the multi-objective problem (MOP); second, some algorithms are used to optimize the objectives; third, it is essential to choose the most appropriate result for the formulated problem, appointed as multicriteria decision making (MCDM) [56]. In time series analysis, the multi-objective approach is used to develop a model that makes predictions with small errors which should vary slightly, in other words, it is adopted to deal with bias–variance trade-off [57] where forecast errors and errors variation are minimized.

According to Marler and Arora [55], in the first step, the MOP is defined, regarding which decision variables, constraints and objectives are stated. The MOP for two J objectives can be stated as follows:

$$\min/\max_{\hat{\theta}} \quad J(\theta) = [J_1(\theta), J_2(\theta)],\tag{5}$$

subject to inequality constraints

$$L \leq \theta_i \leq U,\tag{6}$$

in which $\theta = [\theta_1, \ldots, \theta_i]$, $(i = 1, \ldots, n)$ is a vector of decision variables, L and U are the lower and upper boundaries for each decision variable, and $J_k(\theta)$ is the kth objective to be minimized or maximized.

In this respect, during the MOO step, an optimization algorithm is applied to find the Pareto Front approximation (PF). This set is composed of non-dominated solutions for which there is no other permissible solution that simultaneously improves all the objective functions without sacrificing at least one of the other objective function. Each set of decision variables associated with each element of PF makes up the Pareto Set (PS) [55].

As regards to the MOO approach, most of the algorithms proposed in the literature are based on evolutionary computation and are named multi-objective evolutionary algorithms (MOEA). An evolutionary algorithm is based on nature's laws, and it uses mechanisms such as crossover (an operation used to combine the genetic information of parents to produce a new offspring), mutation (an operator used to maintain the diversity of the population by introducing another level of randomness) and selection (an operation applied to select the fittest individuals according to the objectives to be optimized) [58].

Basically, in an MOEA, the fittest members of a parent in a population will survive and propagate their fitness along with the evolution of its populations of offspring. In the literature, new approaches to deal with MOO are constantly being proposed. Nevertheless, in this paper NSGA-II is adopted [35], considering it is a classical MOO approach and has already shown good results (in terms of accuracy, in general for regression analysis, lower error) to deal with MOP on several fields of knowledge [59,60]. The NSGA-II is a derivative-free method which employs a non-sorting ordering operator and crowding distance criteria in the optimization process. The first operation is used to find a set of equally good solutions closest to the Pareto-optimal front, while the second is employed to promote diversity among members of the solution population. The main idea behind the crowding distance is finding the Euclidean distance between neighboring individuals in the m-dimension space [35,61]. The NSGA-II parameters adopted in this paper are exposed in Section 5, item 4b.

Lastly, in the MCDM step, it is possible to find a preferable set of decision variables (weights in this paper) that allows dealing with the trade-off between the objectives. It means finding the best compromise between the optimized objectives. An approach used for this purpose is TOPSIS, which was proposed by Hwang and Yoon [36]. The TOPSIS was meant to determine the best alternative based on the concepts of the compromised solution. In other words, the chosen solution should be as close to the positive ideal solution as possible, and as far away from the negative ideal solution as possible, considering a measure of similarity. The distance of ideal and negative ideal solutions can be obtained by the Euclidean distance [62,63]. The TOPSIS weights used for each objective are defined in Section 5, item 4c.

5. Methodology

1. Performing EEMD for the datasets and obtaining four IMF and one residue component. The parameters, number of ensemble components, number of components (IMF and residual components) and amplitude of white noise are defined as 100, 5 and 6.4×10^{-7}, respectively. These values are obtained according to Wu and Huang [26]. This paper does not focus on the discussion of the model's performance according to the parameter setting;

2. For each component, the ACF analysis showed, for most cases, that up to four lags are suitable to be used as predictors. Without loss of generality, for all components and states studied, this configuration is used as inputs of each adopted model (BRNN, CUBIST, GBM, PLS, and QRF);
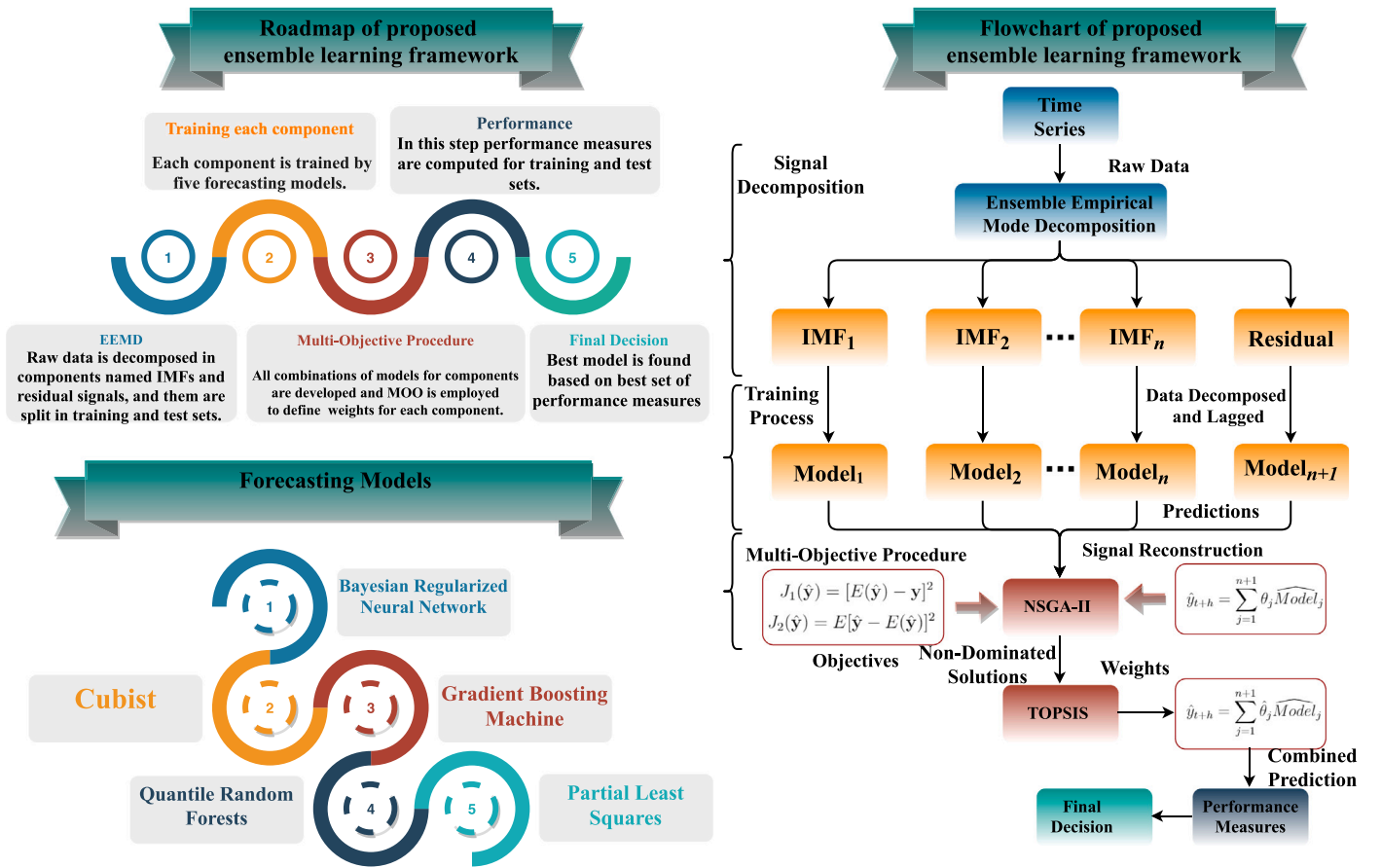
Fig. 3. Flowchart of proposed approach.

3. Training each IMF and residual components using the models mentioned in step 2 by leave-one-out cross-validation with a time slice window (LOOCV-TS), according to

$$y_{(t+1,k)} = f\left\{y_{(t,k)}, y_{(t-1,k)}, y_{(t-2,k)}, y_{(t-3,k)}\right\} + \epsilon \quad \epsilon \sim N(\mathbf{0}, \sigma^2), \quad (7)$$

and forecast the meningitis cases according to the recursive method, as given by Eq. (8),

$$S: \begin{cases} \hat{y}_{(t+h,k)} = f\left\{y_{(t+h-1,k)}, y_{(t+h-2,k)}, \ldots, y_{(t+h-n_y,k)}\right\} & \text{if } h = 1, \\ \hat{y}_{(t+h,k)} = f\left\{\hat{y}_{(t+h-1,k)}, y_{(t+h-2,k)}, \ldots, y_{(t+h-n_y,k)}\right\} & \text{if } h = 2, \\ \hat{y}_{(t+h,k)} = f\left\{\hat{y}_{(t+h-1,k)}, \hat{y}_{(t+h-2,k)}, \ldots, y_{(t+h-n_y,k)}\right\} & \text{if } h = 3, \end{cases}$$
$$(8)$$

in which f is a function related to the adopted model in the training process, $\hat{y}_{(t+h,k)}$ is the forecast value for kth component obtained in the decomposition stage ($k = 1,\ldots,5$) on time t and forecast horizon h ($h = 1, 2, 3$), $y_{(t+h-n_y,k)}$ are the previously notified cases lagged in $n_y = 1,\ldots,4$ and ϵ is the random error which follows a normal distribution (N) with zero mean and constant variance.

The recursive method, also known as an iterated method, can lead to poor accuracy in long forecasting horizons. According to Pouzols and Barros [64], and Veloz et al. [65], the recursive strategy uses forecasting values as a model's input to forecasting the next predicted values. Its main disadvantage is to accumulate the previous forecasting errors in the recursive process. However, the advantage of the recursive method lies in the use of one

model for all processes, i.e., train one model to forecast a one-step-ahead horizon, and use it for multi-step-ahead forecasting task. On the other hand, the direct method uses only past values to predict the future, which is its advantage, as it does not accumulate prediction errors. However, its disadvantage lies in the necessity to fit or training a new model for each forecasting horizon. Therefore, this disadvantage makes the process complex and computationally intensive. Due to separated models are used to forecasting two points, there is no opportunity of handling the dependency between two consecutive predictions.

According to Ma and Fildes [66], comparative studies between Recursive and Direct strategies have shown contradictory results as to which strategy achieves better forecasting performances than others. While Rana and Rahman [67] achieved better results for direct method over the recursive method, Xue et al. [68] provided evidence of better results from the recursive method over the direct method. There is a trade-off between computational cost and accuracy in the choice of a multi-step-ahead forecasting method.

Machine learning models, sometimes, have high training time, either due to the use of different training strategies such as cross-validation (k-fold or LOOCV-TS), or even due to the number of parameters to be tuned. In this context, because in this paper several models are evaluated, to find an efficient ensemble learning forecasting model to study meningitis cases, the recursive forecasting strategy is adopted. Moreover, the forecasting horizons are defined as one, two, and three months-ahead, which are considered short-term. Therefore, even though the recursive method could lead to high forecasting errors, it is used in this

paper due to its lower computational cost. However, the direct method can be considered for this study.

In this paper, the hyperparameters of the adopted models in step 2 are obtained during the LOOCV-TS by a grid-search (GS) and each situation is presented in Table B.1 of Appendix B. Giving a set of combinations of the hyperparameters values, GS consists of searching exhaustively throughout the combinations which is the best set of hyperparameters that minimize/maximize any criteria; in this paper, root mean squared error is considered. The LOOCV-TS consists of using an initial slice of training data as an estimation set, and then it validates the performance of the trained model over a desired number of horizons. The next step consists of using the data for validation while incorporating it into the training set, and the process is performed until the entire training set is used. The model performance is obtained as the average criteria used in each iteration [69].

4. To aggregate the EEMD's components, the forecasting obtained in step 3 are used. To choose the order of the models used to train and predict the results of each component, a set of 3125 models is evaluated. These models are obtained by permuting the five forecasting models QRF, PLS, BRNN, GBM, and CUBIST (in this order) for the five EEMD components. Therefore, the methodology used to select the models' order is the grid-search. Table 1 shows a sample of 3 out of 3125 ensemble learning models, randomly selected, where the order of models for each component is detailed.

After defining the weights for each model-component, suppose that the ensemble learning model number 295 has better accuracy for out-of-sample forecasting, then the order of the models used for each component is the presented in Table 1.

For each combination, the optimization process is conducted according to what is described as follows:

(a) In the MOP, the cost function, for each combination of models and components, is stated as follows:

$$\hat{y}_{t+h} = \sum_{j=1}^{n=4} \theta_j \widehat{IMF}_{Model_j} + \theta_5 \widehat{Residual}_{Model_5}, \quad (9)$$

in which $\hat{y}_{t+h}$ is the predicted value on time t and h horizon, $\widehat{IMF}_{Model_j}$ ($j = 1, \ldots, 4$) and $\widehat{Residual}_{Model_5}$ are the predictions of each component and $\theta = [\theta_1, \ldots, \theta_j]$ is the weight vector to be estimated, in which $\theta_j \sim Unif[-2, 2]$ [70]. Considering the bias–variance framework [57], the objectives are defined as follows:

$$J_1(\hat{\mathbf{y}}) = \overbrace{[E(\hat{\mathbf{y}}) - \mathbf{y}]^2}^{Bias-Error} \quad \text{and} \quad J_2(\hat{\mathbf{y}}) = \overbrace{E[\hat{\mathbf{y}} - E(\hat{\mathbf{y}})]^2}^{Variance-Stability}, \quad (10)$$

$$E(\hat{\mathbf{y}}) = \frac{1}{n} \sum_{i=1}^{n} \hat{y}_i \quad \text{and} \quad \mathbf{y} = \frac{1}{n} \sum_{i=1}^{n} y_i, \quad (11)$$

in which $\hat{y}_i$ is the ith predicted value through the aggregation of components and y_i is the ith observed value. The error is computed through the mean squared error criterion, and the stability by the variance of the forecasting errors.

(b) In the sequence, the NSGA-II is applied (one run) and the PF approximation is obtained. In this paper, the parameters used for this algorithm, population size (number of candidate solutions), maximum number of generations (stopping criterion), crossover and mutation probability, crossover and mutation distribution, are defined as 100, 100, 0.9, 0.1, 0.7 and 0.2, respectively. The crossover rate is set to 0.9 because it allows new structures to be introduced into the population at a faster rate, whereas

the mutation rate is set to 0.1 because it prevents a given position from becoming stationary in a set of values for the parameters to be optimized [35];

(c) To find the best set of weights, between all candidates, the TOPSIS approach is employed, in which the weights for bias (error) and variance objectives are 50% and 50%, respectively;

5. Forecasting the meningitis cases, out-of-sample, according to

$$\hat{y}_{t+h} = \sum_{j=1}^{n=4} \hat{\theta}_j \widehat{IMF}_{Model_j} + \hat{\theta}_5 \widehat{Residual}_{Model_5}, \quad (12)$$

where $\hat{\theta}_j$ is the estimated weight;

6. Computing the performance measures MAE, RRMSE and sMAPE given by,

$$MAE = \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{n} \right|, \quad (13)$$

$$sMAPE = \frac{2}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{|y_i| + |\hat{y}_i|}, \quad (14)$$

$$RRMSE = 100 \times \frac{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}}{\frac{1}{n} \sum_{i=1}^{n} y_i}, \quad (15)$$

where n represents the number of observations, y_i and $\hat{y}_i$ are the ith observed and predicted values, respectively. The best model is the one with the best capacity of out-of-sample generalization in terms of MAE, RRMSE, and sMAPE.

Aiming to compare the forecast errors of two models, the Diebold–Mariano (DM) test [71] is applied. In this paper, the lower tail priori hypothesis H is given by Eq. (16),

$$H : \begin{cases} H_0 : \epsilon_i^P = \epsilon_i^C \\ H_1 : \epsilon_i^P < \epsilon_i^C, \end{cases} \quad (16)$$

and statistic of DM test is given by Eq. (17),

$$DM = \frac{\frac{\sum_{i=1}^{i=1} \left[L(\epsilon_i^P) - L(\epsilon_i^C) \right]}{n}}{\sqrt{\frac{S^2}{n}}} s^2, \quad (17)$$

in which L is a loss function that can estimate the accuracy of each model, ϵ_i^P is the error of the proposed model, ϵ_i^C is the error of the compared model, and s^2 is an estimate for the variance of $d_i = L(\epsilon_i^P) - L(\epsilon_i^C)$. By using the hypothesis defined for the DM test, the interest lies in knowing if the errors for the proposed model are lower than the compared model. If the null hypothesis is rejected, it is possible to say that there is statistically evidence that there is a reduction in the errors of the EEMD-HTE-MOO model regarding the compared model at the α level of significance.

In addition, Fig. 3 presents the modeling process.

The results presented in Section 6 are generated using the processor Intel(R) Core(TM) i5-4200U central processing unit of 1.6 Hz in

Table 1

Randomly selected ensembles learning models and their order with respect to the EEMD components.

Grid-search index	IMF_1	IMF_2	IMF_3	IMF_4	Residual
295	CUBIST	GBM	PLS	BRNN	QRF
1831	QRF	PLS	GBM	CUBIST	BRNN
2639	GBM	BRNN	QRF	PLS	CUBIST

Windows 10 operating system with 8GB of random access memory. The R software [72] is adopted to perform the modeling. The packages hht [73], caret [74], mco [75] and MCDM [76] are used on the steps 1, 3, 4b and 4c, respectively. Moreover, the codes adopted in the proposed methodology are available at https://github.com/MRibeiro2107/-JBI-19-1008.

6. Results and discussions

This section presents the results and discussions of the information obtained from the developed experiments. Table B.1 contained in Appendix B presented the control hyperparameters of the adopted models for the task of training and predicting each component of the decomposition stage, as well as the data trained without EEMD decomposition. In the context of the structure of the hybrid, the framework proposed Table C.1 contained in Appendix C shows the set of weights, and the respective models used to train and predict each IMF and residue. Considering all the forecast horizons, for the developed models, the GBM model is the most used to train the components, while QRF is the least employed for this task. In both cases, GBM and QRF are ensemble learning approaches; however, the first model is based on boosting, while the second one on bagging. The objective of Sections 6.1–6.3 is to present the main results obtained with the proposed approach against various classes of models. Faced with this, three different comparisons were developed and performance on the out-of-sample forecast (for test set) is presented. Through these comparisons, the prediction accuracy of the proposed framework is effectively evaluated. In Tables 2 to 4, the best results are presented in bold.

6.1. Comparisons I: Evaluation of proposed ensemble learning and decomposed homogeneous optimized ensemble learning models

The comparisons I is designed to verify the forecast performance of the proposed hybrid framework by comparing it with five models based on decomposition, which uses the same model for all components and MOO in the aggregation step, namely EEMD-BRNN-MOO, EEMD-CUBIST-MOO, EEMD-GBM-MOO, EEMD-QRF-MOO, and EEMD-PLS-MOO. The comparisons are shown in Table 2 and additional discussion are also presented. These results shows that the approach considered as a heterogeneous ensemble of components is more accurate than other approaches that consider the same model for all IMF and residual, to forecast the number of meningitis cases one, two and three-month-ahead of time. According to Zhang et al. [77], this effectiveness is associated with the diversity used by the heterogeneous ensemble approach, which is an efficient and simple way to perfect forecasting accuracy and stability (lower standard deviation of the errors). Faced with this, it makes the predictive model more robust.

Concerning the one-month-ahead forecast, for all states, the proposed framework achieves better accuracy than the compared models. Regarding MAE criterion, the compared models increasing the forecasting errors regarding proposed methodology, which ranges between 22.73% and 36.36%, 36.87% and 138.22%, 0.72%, and 54.35%, as well as 41.51% and 66.04% for the Brazilian states MG, SP, PR, and RJ, respectively. Regarding the sMAPE the compared models increasing the forecasting errors regarding proposed ensemble learning model, which is ranged between 1.81%-3.51%, 4.72%–17.72%, 0.46%–5.96% and 2.16%–3.32% for the states of MG, SP, PR, and RJ, respectively. Considering the RRMSE, the compared models increasing the forecasting errors regarding proposed ensemble learning model, which is ranged between 2.09%–3.98%, 3.76%–15.75%, 0.74%–6.58% and 2.22%–4.75% for the states of MG, SP, PR, and RJ, respectively. The same behavior is reproduced when the forecasting horizon is two and three-month-ahead of time and in some situations with a greater magnitude of the reduction.

In this scenario, the largest and smallest reductions in these measures occurred when the proposed approach and EEMD-QRF-MOO for

the states of SP and PR are compared. This is associated with the fact that in the proposed framework, for these states, there is the use of two and four different models, respectively, to compose the heterogeneous ensemble approach. This diversity allows the developed model to achieve the best performance, regarding RRMSE and sMAPE, associated to the fact that the manipulation of some algorithms to build the ensemble is an important strategy to get an efficient model [78]. In this respect, each model learns different patterns of the data, and when the results are combined, they build an effective model. These results corroborate with the findings of Heinermann and Kramer [79],Xiang et al. [80], proving that using different models for the decomposed components is promising for time series forecasting.

6.2. Comparisons II: Evaluation of proposed ensemble learning and non-decomposed models

The comparison II is designed to verify the forecasting performance of the proposed hybrid framework by comparing it with five models which do not consider EEMD for signal decomposition, namely BRNN, CUBIST, GBM, QRF, and PLS. The comparisons are shown in Table 3 and additional discussion is presented. Regarding these results, it is observed that the use of signal decomposition, specifically in the case of the EEMD approach, can enhance the model's performance. This shows that the EEMD approach is suitable for decomposing the time series of meningitis case for all states. For this reason, the idea of decomposition and ensemble is feasible, and the proposed framework can overcome the drawbacks of individual models by removing the noise of the original data.

Considering the average of each criterion for each forecast horizons, there is a reduction in the criteria for the proposed approach compared to models that do not consider signal decomposition. The increasing on MAE criterion compared models regarding proposed ensemble learning model is ranged between 95.03%–116.18%, 64.90%–91.48%, 75.20%–99.88%, 98.52%–125.62% for the states of MG, SP, PR, and RJ. In its turn, for sMAPE and RRMSE, the same behavior is observed. The smallest reduction is visualized by comparing the proposed framework to the PLS model, and the largest when it is performed comparison with QRF, for both criteria.

The results of one-month-ahead up to three-month-ahead forecast for meningitis cases were obtained. The comparisons between the proposed approach and the models without decomposition showing that the performance measures of the proposed model are lower than others obtained from the classic single models. For this reason, the proposed model is more adequate than the single models without decomposition to forecasting meningitis cases. The EEMD presents itself as being the more robust and accurate model, after the nonlinear and non-stationary signals are separated into a series of components. Consequently, this effectiveness allows an improvement of the model's performance by removing the noise from the time-series [81,82].

6.3. Comparisons III: Evaluation of proposed ensemble learning and decomposed heterogeneous direct integrated ensemble learning models

The comparison III is designed to verify the forecasting performance of the proposed hybrid framework by comparing it with the same structure of EEMD-HTE which uses DI to aggregate the EEMD components, namely EEMD-HTE-DI. The comparisons are shown in Table 4, and additional discussion is presented.

In accordance, with these results, it can be stated that the use of MOO in the WI aggregation allows an improvement of the model's accuracy. Based on the evaluation of all criteria for the EEMD-HTE-MOO, its approach achieves more optimum accuracy than the EEMD-HTE-DI approach. Thus, the proposed methodology can generate robust models that have enhanced forecasting performance with this model combined with it based on decomposition, ensemble, and MOO.

Table 2
Performance measures of proposed and decomposed homogeneous optimized ensemble learning models.

State	Model	Forecasting Horizon								
		One-month-ahead			Two-months-ahead			Three-months-ahead		
		MAE	sMAPE	RRMSE	MAE	sMAPE	RRMSE	MAE	sMAPE	RRMSE
	EEMD-MOO-BRNN	10	11.34%	14.04%	12.58	14.38%	15.79%	11.83	13.62%	14.75%
	EEMD-MOO-CUBIST	9.08	10.39%	12.93%	8.75	10.14%	13.13%	10.75	12.37%	14.30%
MG	EEMD-MOO-GBM	9	10.48%	13.36%	9.67	11.03%	13.84%	7.83	8.77%	11.75%
	EEMD-MOO-PLS	10.50	11.86%	14.81%	13.25	15.14%	16.94%	12.33	14.17%	15.68%
	EEMD-MOO-QRF	9.17	10.16%	13.96%	12.50	14.37%	17.31%	11.67	13.20%	16.04%
	Proposed	**7.33**	**8.35%**	**10.83%**	**7.92**	**9.02%**	**10.63%**	**6.67**	**7.37%**	**10.38%**
	EEMD-MOO-BRNN	151.42	26.86%	28.01%	104.92	16.86%	21.55%	137.33	22.76%	27.24%
	EEMD-MOO-CUBIST	92.50	14.62%	18.89%	103.83	16.44%	21.64%	115.83	18.44%	23.78%
SP	EEMD-MOO-GBM	146.50	23.52%	30.02%	127.92	19.60%	26.74%	161.75	26.12%	34.24%
	EEMD-MOO-PLS	106.50	17.02%	21.72%	157.50	27.18%	28.56%	172.58	28.96%	33.65%
	EEMD-MOO-QRF	161	27.62%	30.89%	127.67	20.43%	24.67%	182.58	31.73%	32.35%
	Proposed	**67.58**	**9.90%**	**15.14%**	**80.67**	**12.02%**	**17.04%**	**87.08**	**12.77%**	**17.21%**
	EEMD-MOO-BRNN	15.42	10.95%	14.24%	16.33	11.41%	15.06%	33.83	28.72%	28.43%
	EEMD-MOO-CUBIST	15.83	11.19%	15.18%	27.17	21.45%	25.55%	15.25	10.86%	14.67%
PR	EEMD-MOO-GBM	17.75	12.41%	17.52%	38.83	33.13%	34.74%	47	43.93%	39.73%
	EEMD-MOO-PLS	15.83	10.77%	15.14%	35.50	30.08%	31.50%	40.17	35.06%	33.82%
	EEMD-MOO-QRF	11.58	8.03%	12.84%	15.33	10.37%	19.80%	17.50	11.52%	17.42%
	Proposed	**11.50**	**7.57%**	**12.10%**	**12.50**	**8.61%**	**11.95%**	**11.75**	**7.70%**	**13.40%**
	EEMD-MOO-BRNN	6.33	7.47%	8.57%	7.08	7.99%	9.12%	4.83	5.46%	7.21%
	EEMD-MOO-CUBIST	7.33	8.71%	9.71%	8.25	9.36%	10.58%	10.75	12.64%	13.91%
RJ	EEMD-MOO-GBM	6.25	7.13%	8.11%	8.58	9.73%	11.06%	6.92	7.89%	8.43%
	EEMD-MOO-PLS	7.33	8.29%	9.32%	7.42	8.32%	9.19%	6	6.76%	7.99%
	EEMD-MOO-QRF	6.75	8.13%	10.64%	8.42	9.87%	12.49%	8	9.20%	10.21%
	Proposed	**4.42**	**4.98%**	**5.89%**	**6.58**	**7.37%**	**8.52%**	**4.08**	**4.52%**	**6.49%**

Table 3
Performance measures for proposed and non-decomposed models.

State	Model	Forecasting Horizon								
		One-month-ahead			Two-months-ahead			Three-months-ahead		
		MAE	sMAPE	RRMSE	MAE	sMAPE	RRMSE	MAE	sMAPE	RRMSE
	BRNN	15.47	17.56%	19.38%	13.83	15.79%	18.14%	13.67	15.60%	17.66%
	CUBIST	14.99	17.05%	20.14%	14.67	16.73%	20.12%	13.08	14.96%	18.57%
MG	GBM	16.97	19.21%	21.88%	14.67	16.72%	19.53%	15.74	17.88%	20.99%
	PLS	15.72	17.82%	19.65%	13.92	15.89%	18.18%	13.93	15.88%	17.99%
	QRF	15.58	17.60%	20.05%	15.33	17.30%	19.50%	12.67	14.48%	16.93%
	Proposed	**7.33**	**8.35%**	**10.83%**	**7.92**	**9.02%**	**10.63%**	**6.67**	**7.37%**	**10.38%**
	BRNN	120.37	17.99%	24.60%	135.33	20.39%	27.14%	160.56	23.82%	30.24%
	CUBIST	122.94	18.62%	23.92%	151.33	22.79%	28.65%	153.82	22.93%	27.96%
SP	GBM	117.98	17.86%	23.37%	143.83	21.26%	29.62%	178.91	26.14%	32.15%
	PLS	110.80	16.76%	22.53%	117	17.97%	23.41%	160.26	24.12%	28.83%
	QRF	129.83	19.12%	26.82%	148.83	21.97%	30.39%	171.96	25.69%	32.34%
	Proposed	**67.58**	**9.90%**	**15.14%**	**80.67**	**12.02%**	**17.04%**	**87.08**	**12.77%**	**17.21%**
	BRNN	20.36	14.49%	22.52%	22.58	16.24%	25.05%	23.21	16.38%	24.29%
	CUBIST	20.58	14.91%	24.88%	22	16.08%	28.90%	25.42	18.08%	26.75%
PR	GBM	21.41	15.15%	25.21%	23.67	17.14%	27.91%	25.76	18.39%	27.87%
	PLS	19.13	13.71%	22.22%	20.58	14.82%	24.06%	22.92	16.15%	24.18%
	QRF	22.17	15.74%	23.75%	24.67	17.85%	28.19%	25.75	18.40%	25.89%
	Proposed	**11.50**	**7.57%**	**12.10%**	**12.50**	**8.61%**	**11.95%**	**11.75**	**7.70%**	**13.40%**
	BRNN	11.79	13.25%	15.46%	11.58	13.06%	15.98%	10.66	12.09%	14.75%
	CUBIST	11.60	12.99%	15.26%	11.42	12.87%	15.90%	10.86	12.29%	15.24%
RJ	GBM	10.45	11.81%	14.34%	9.50	10.80%	13.39%	9.99	11.36%	14.00%
	PLS	10.50	11.83%	13.93%	10.50	11.83%	14.49%	10.11	11.42%	14.24%
	QRF	11.96	13.40%	14.83%	12	13.59%	15.93%	11.50	13.04%	15.43%
	Proposed	**4.42**	**4.98%**	**5.89%**	**6.58**	**7.37%**	**8.52%**	**4.08**	**4.52%**	**6.49%**

The EEMD-HTE-MOO presents a better generalization capacity than the EEMD-HTE-DI, considering that the nonlinearities are accommodated by the use of different weights for each component. In general, this shows that the proposed ensemble model can combine the advantages of the single forecasting models, through the use of MOO, making it possible to add more accuracy and stability to the model [83]. Also, the effectiveness of the proposed approach, according to Bui et al. [84], can be attributed to the process of construction of the combined model using the optimization. Consequently, the MOO allows dealing with the trade-off between bias–variances while assigning different weights for each model. By minimizing the forecast and standard deviation errors

it becomes possible to obtain an efficient model with both accuracy and stability improved simultaneously presenting a balance between the two optimized objectives.

As presented in Section 5, the results presented in Tables 2 to 4 are obtained through the recursive method to forecasting multi-step-ahead meningitis cases. Through this method, the errors can be accumulated according to the growth of the forecasting horizon. This phenomenon can be observed in MG, PR, and RJ states when are compared the first two forecasting horizons (regarding to MAE and sMAPE), and for SP state in the three forecasting horizons. However, in some cases, the proposed model reduces the forecasting errors, even

Table 4
Performance measures for proposed and decomposed heterogeneous direct integrated ensemble learning models.

State	Model	Forecasting Horizon								
		One-month-ahead			Two-months-ahead			Three-months-ahead		
		MAE	sMAPE	RRMSE	MAE	sMAPE	RRMSE	MAE	sMAPE	RRMSE
MG	**Proposed**	**7.33**	**8.35%**	**10.83%**	**7.92**	**9.02%**	**10.63%**	**6.67**	**7.37%**	**10.38%**
	EEMD-HTE-DI	8.92	9.93%	12.51%	8.75	10.14%	13.13%	8.58	9.43%	12.51%
SP	**Proposed**	**67.58**	**9.90%**	**15.14%**	**80.67**	**12.02%**	**17.04%**	**87.08**	**12.77%**	**17.21%**
	EEMD-HTE-DI	101.92	15.31%	20.18%	102.25	15.49%	20.52%	132.83	20.57%	25.87%
PR	**Proposed**	**11.50**	**7.57%**	**12.10%**	**12.50**	**8.61%**	**11.95%**	**11.75**	**7.70%**	**13.40%**
	EEMD-HTE-DI	12.67	8.73%	13.43%	16	10.81%	19.85%	17.75	11.68%	17.48%
RJ	**Proposed**	**4.42**	**4.98%**	**5.89%**	**6.58**	**7.37%**	**8.52%**	**4.08**	**4.52%**	**6.49%**
	EEMD-HTE-DI	5.92	6.72%	8.12%	7.83	8.68%	10.51%	5.83	6.65%	8.05%

Table 5
Average standard deviation of errors obtained by each model in forecast out-of-sample (test set forecast).

Model	MG	SP	PR	RJ	Model	MG	SP	PR	RJ
Proposed	**5.84**	**96.97**	**17.68**	**6.36**	EEMD-MOO-QRF	9.06	126.69	23.41	9.72
EEMD-MOO-BRNN	6.29	104.19	20.40	7.45	GBM	12.75	189.43	36.73	12.79
EEMD-MOO-CUBIST	6.36	97.25	21.41	7.96	PLS	12.96	166.82	32.62	12.81
EEMD-MOO-PLS	7.41	113.27	23.99	8.10	BRNN	13.67	182.64	33.24	14.05
EEMD-HTE-DI	7.47	124.95	23.51	8.16	CUBIST	13.97	177.90	36.55	13.98
EEMD-MOO-GBM	7.82	145.54	27.33	8.24	QRF	14.29	198.99	34.66	14.23

with the increase of the forecasting horizon. Nevertheless, according to Taieb et al. [85], there is greater evidence of errors accumulation in long-term forecasting horizons, once the input vector adopted by the predictive model can be formed only by previously forecasting, instead of observed values. However, this is not the case of this paper, once the forecast horizon is short (h equals to 1, 2, and 3). In its turn, as three values (n_y=4) are adopted as input to the models, there are always n_y-h+1 observed values in the input vector. Therefore, the forecasting errors can reduce according to the adopted forecasting horizons.

Additionally, this information is supported by the results presented in Table 5, in which the error's standard deviation is presented for each model throughout the twelve study scenarios. The best results are presented in bold.

By considering the behavior of the meningitis monthly incidence, there is no clear trend in the time series. In this respect, once the forecasting value of two months-ahead is used as models' input to forecasting three-months-ahead, the information provided in the previous forecasting value can be beneficial to the next forecasting, even though they are different of observed values. In this context, the forecasting errors of the next step-ahead can be reduced, which is observed in Tables 2 to 4. The same is observed in the results achieved by Xiong et al. [86] when the hybrid model which combines seasonal and trend decomposition using locally estimated scatterplot smoothing with extreme learning machine was applied to forecast the vegetable market, and Moreno et al. [87] when the authors proposed a hybrid model based on singular spectral analysis, variational mode decomposition, and long short-term memory to forecasting multi-step-ahead wind speed.

6.4. Statistical tests and graphical analysis.

In this subsection, the results of DM tests are used to evaluate the forecasting validity of the proposed model, further verifying the performance of this framework. The EEMD-HTE-MOO errors are compared to the models presented in Sections 6.1–6.3 , and the results of DM test statistics for the errors during the task of forecasting the out-of-sample meningitis cases are presented in Table 6.

The comparisons are developed according to three situations. First, the proposed framework and the structures that use a homogeneous ensemble of components and MOO are compared. According to this, for the one-month-ahead forecast, the results of Table 6 highlights that in 95% of the comparisons, the errors of the proposed approach are

statistically smaller than errors of the compared approaches. In the cases of EEMD-QRF-MOO and EEMD-GBM-MOO, for the states of PR and RJ respectively, the proposed approach presents only smaller errors than these approaches. Considering two-month-ahead or three-month-ahead forecasting, in 85% and 95% of the comparisons, respectively, there is a statistical difference between the errors. For this scenario, the biggest differences are found when the proposed model and EEMD-PLS-MOO approaches are compared with two and three-month-ahead forecast in the states of MG and PR, respectively. It is possible to perceive that the proposed approach shows better performance in a statistical context, in 91.67% of the cases, allowing an inference that makes the developed framework efficient for the proposed task. Second, when the errors of the developed structure and the models that do not employ signal decomposition are compared, for all cases, there is a statistical difference. By comparing these results, Table 6 illustrates that the biggest differences are −3.82, −37.76 and −3.69 in MG and SP, states on the task of one, two and three-month-ahead forecast, respectively, for models GBM and QRF. Additionally, the smallest differences are −1.24, −1.24 and −1.40 for the state of PR, for the three forecast horizons, for models PLS, QRF, and BRNN, respectively. Therefore, these results showed that the use of signal decomposition to forecasting meningitis cases allows the enhancement of the models' accuracy. Also, the results presented in Table 6, by comparing the EEMD-HTE-MOO and EEMD-HTE-DI, in 86.67% of the cases, there is a statistical difference between the errors. Moreover, on the task of forecasting the meningitis cases in one and two-month-ahead for the state of PR, and three-month-ahead for the state of RJ, there is no statistical difference between the errors. Moreover, the smallest and biggest differences are observed on three-month-ahead for the state of MG (−1.32) and two-month-ahead forecast for the state of RJ (−1.92), respectively. Regarding what has previously been mentioned, it can be stated that the proposed approach showed a significant improvement concerning the use of the DI strategy for the task of grouping the EEMD components.

Therefore, regarding the results presented in Sections 6.1–6.4 , the proposed approach showed better accuracy than the compared models. In parallel, the proposed framework achieves excellent performance in 83.33% and 25% of the cases, and good results in 16.67% and 25% of the cases for sMAPE and RRMSE, respectively Li et al. [101]. This shows that the use of signal decomposition, heterogeneous ensemble components and MOO combined can enhance the model's performance, taking into account the combination of each methodology's expertise.

Table 6
Diebold–Mariano test results.

Model	MG			SP		
	One-month-ahead	Two-months-ahead	Three-months-ahead	One-month-ahead	Two-months-ahead	Three-months-ahead
EEMD-MOO-QRF	−1.32*	−2.91***	−2.48**	−2.88***	−2.06**	−3.05***
EEMD-MOO-PLS	−3.08***	−7.02***	−2.97***	−2.91***	−3.57***	−4.23***
EEMD-MOO-BRNN	−2.12**	−4.07***	−1.65*	−3.77***	−2.18**	−2.78**
EEMD-MOO-GBM	−2.02**	−1.58*	−1.22*	−2.62**	−1.90**	−2.43**
EEMD-MOO-CUBIST	−1.99**	−1.35*	−2.55**	−2.63**	−2.08**	−2.37**
QRF	−3.62***	−2.84***	−2.11**	−1.96**	−2.39**	−1.96**
PLS	−3.51***	−4.52***	−2.40**	−2.17**	−2.07**	−3.48***
BRNN	−3.54***	−5.99***	−2.42**	−2.08**	−2.23**	−1.87**
GBM	−3.82***	−2.71***	−2.03**	−1.83**	−37.76***	−2.01**
CUBIST	−2.81***	−2.41***	−2.02**	−2.09**	−2.65**	−2.48**

Model	PR			RJ		
	One-month-ahead	Two-months-ahead	Three-months-ahead	One-month-ahead	Two-months-ahead	Three-months-ahead
EEMD-MOO-QRF	−0.58	−1.07	−1.39*	−1.55*	−1.84**	−1.72**
EEMD-MOO-PLS	−2.26**	−2.95***	−12.61***	−2.50**	−0.59	−2.94***
EEMD-MOO-BRNN	−2.43**	−1.91**	−6.89***	−2.16*	−4.00***	−1.56*
EEMD-MOO-GBM	−1.67*	−2.35**	−2.74***	−2.96***	−1.10*	−0.95
EEMD-MOO-CUBIST	−1.59*	−2.64**	−0.80	−2.35**	−2.08**	−3.76***
QRF	−1.31*	−1.24*	−1.91**	−2.97***	−2.35**	−3.69***
PLS	−1.24*	−1.28*	−1.40*	−2.41**	−1.45*	−1.51*
BRNN	−1.29*	−1.30*	−1.40*	−3.02***	−1.81**	−2.56**
GBM	−1.30*	−1.33*	−1.55*	−2.12**	−1.43*	−5.19***
CUBIST	−1.33*	−1.26*	−1.42*	−3.14***	−1.65*	−2.13**

***1% significance level.

**5% significance level.

*10% significance level.

(a) Pareto Front   (b) Observed and Predicted values   (c) ACF of the residuals
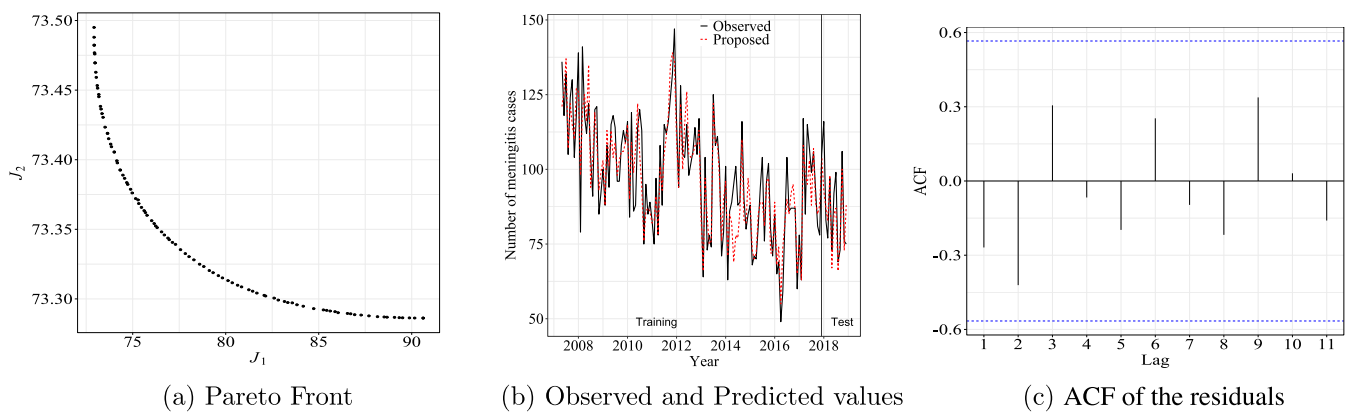
Fig. 4. Performance analysis of proposed framework in MG state for one-month-ahead forecast.

(a) Pareto Front   (b) Observed and Predicted values   (c) ACF of the residuals

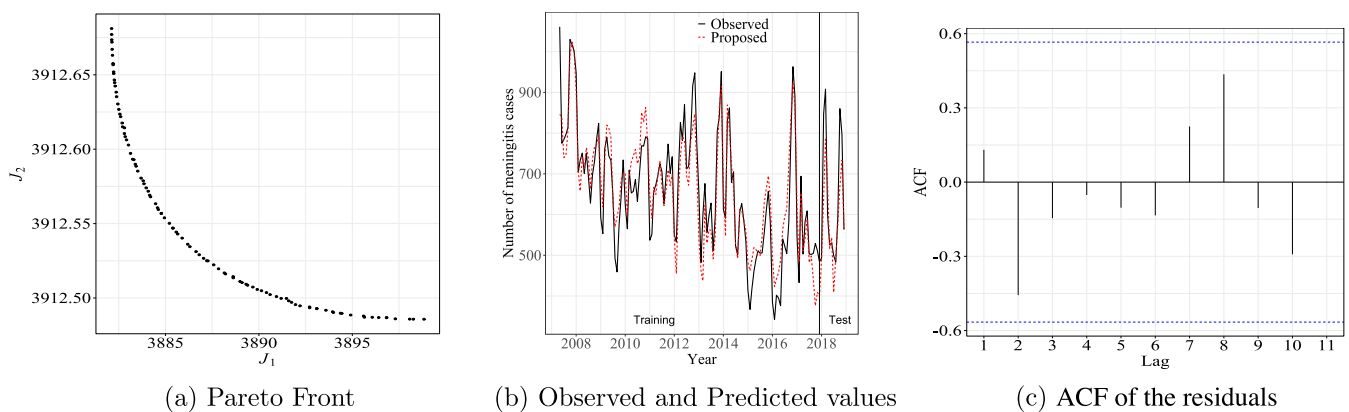**Fig. 5.** Performance analysis of proposed framework in SP state for one-month-ahead forecast.

In addition to the presented analysis, Figs. 4–7 illustrate the Pareto Front (a) of the MOO step, the observed and predicted time series (b) and the residual autocorrelation function of the trained models (c) for the task of one-month-ahead forecasting for the states of MG, SP, PR, and RJ, respectively.
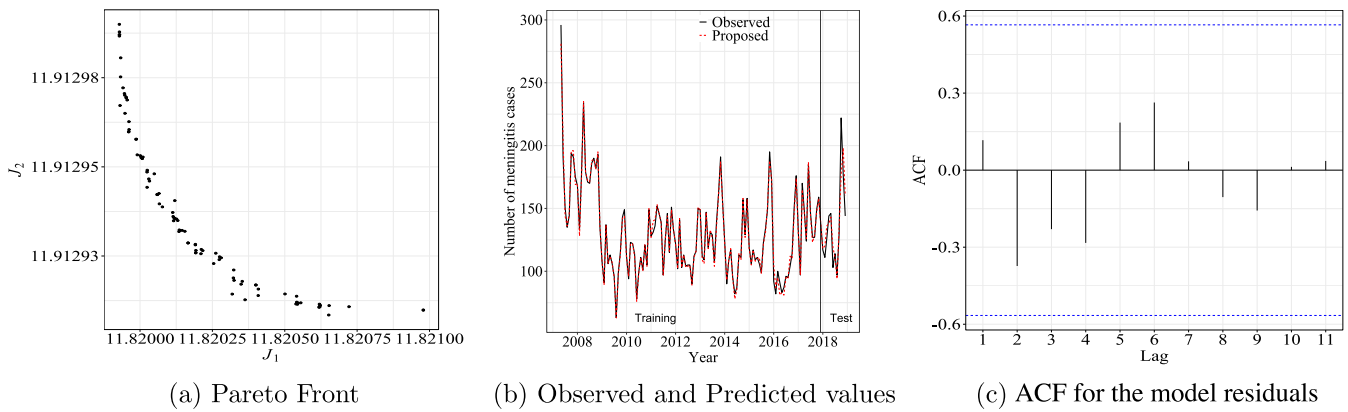
(a) Pareto Front   (b) Observed and Predicted values   (c) ACF for the model residuals

Fig. 6. Performance analysis of proposed framework in PR state for one-month-ahead forecast.

(a) Pareto Front   (b) Observed and Predicted values   (c) ACF of the residuals
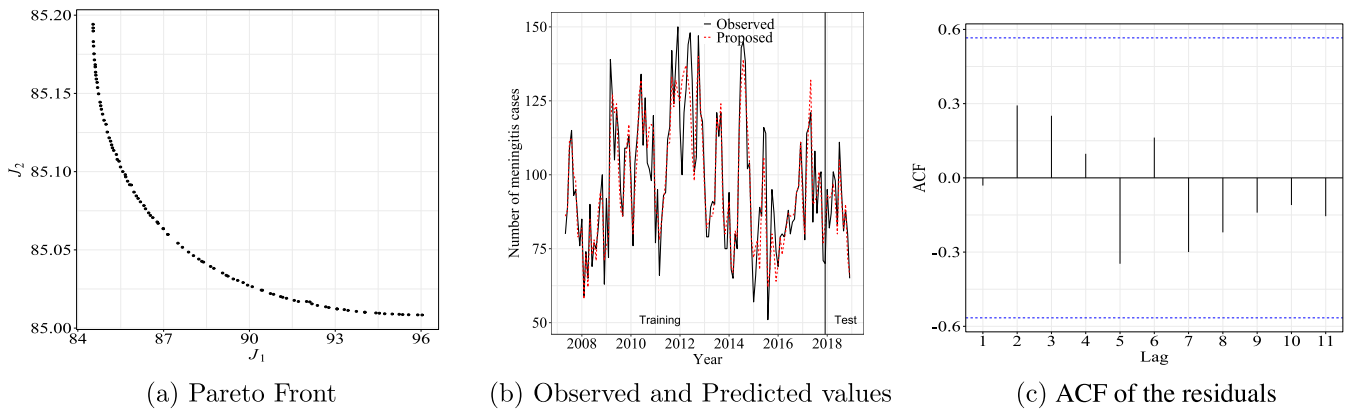
Fig. 7. Performance analysis of proposed framework in RJ state for one-month-ahead forecast.

Considering what is shown by Figs. 4(a), 5(a), 6(a) and 7(a) evidence of a trade-off among the objectives adopted on MOO can be seen, in other words, depending on the weights used to create the ensemble obtained from the EEMD, there is a bias increase while the variance decreases. In this aspect, the use of MOO is adequate, because it allows the obtaining of an efficient model that is able to reach small forecast errors and lower standard deviation errors. The same behavior is replicated for the other two forecast horizons.

During this round, Figs. 4(b), 5(b), 6(b) and 7(b) show that the data behavior is learned by the models, in most of the cases, which allows predictions compatible with the observed values. That is, the meningitis cases forecasted are close to the observed values. The good performance regarding accuracy obtained in the training phase persists in the test stage, indicating that the hybrid framework is robust to reach the developed predictions.

The overfitting phenomenon occurs when the model has great generalization in the training set, but not in test set or out-of-sample forecasting. To avoid it, two approaches were considered. First, each adopted model was trained using a cross-validation procedure, as described in the methodology section, to prevent overfitting. And second, when the bias and variance are objectives adopted in multi-objective optimization, the trade-off between these measures is considered which leads to overfitting treatment. Also, by the illustrated trough of predicted and observed values (Figs. 4(b), 5(b), 6(b) and 7(b)), once a similar performance is observed in training and test sets, there is no evidence of overfitting.

Finally, according to what is shown in Figs. 4(c), 5(c), 6(c) and 7(c), it is possible to evaluate the residuals (autocorrelation function) of final models adopted in each case study, from training the set results. The Box–Ljung test is applied to the residuals from the final ensemble

models to determine whether residuals are random. In both cases, this test shows that the first 11 lag autocorrelation among the residuals are zero ($\chi^2_{11} = 0.84$ - 9.28, p-value >0.05), indicating that the residuals are random and that the model provides an adequate fit to the data [39]. These results are supported by the Figures of residuals ACF since all lags are between confidence interval extremes.

7. Conclusion and future works

In this paper, a hybrid framework was proposed which combines signal decomposition, a heterogeneous ensemble of EEMD components and WI strategy do aggregate the decomposed components. The MOO approach used to find the PF and TOPSIS is employed to find the best set of weights used during the WI stage. This approach is applied to forecasting three different horizons (one, two, and three-month-ahead) of the meningitis cases for the states of MG, SP, PR, and RJ, Brazil. The performance of the proposed framework was compared to the performance of the models that do not employ signal decomposition, models that apply the same model to forecast all components obtained in the decomposition stage combined with MOO to signal reconstruction, and finally with an approach that employs signal decomposition, heterogeneous ensemble of components and DI strategy to reconstruct the original signal. The MAE, sMAPE, and RRMSE criteria, as well as the DM test, were adopted to evaluate the performance of the developed approach. Finally, a residual analysis was conducted by the ACF plot and Ljung–Box [39] test to validate the proposed hybrid model.

According to the forecasting results, it is possible to conclude that:

1. Employing signal decomposition improves the final results on not applying decomposition;

Table A.1
Related works to machine learning and general models for the epidemiologic time series forecasting.

Publication	Disease	Models	Objective	Finding
Ch et al. [88]	Malaria	SVR-FFA, ANNs and ARMA	Predicting the malaria incidences in India using SVR and FFA.	Proposed approach outperform compared models.
Guo et al. [89]	Dengue	SVR, GBM,LASSO and GAM	Use the state-of-the-art machine learning algorithms to develop dengue forecasting in China.	The SVR model had the consistently smallest prediction error rates for tracking the dynamics of dengue and forecasting the outbreaks.
Scavuzzo et al. [90]	Dengue	SVR, ANNs, KNN and DT	Temporal modeling of the oviposition activity (measured weekly on 50 ovitraps in a north Argentinean city) of *Aedes aegypti (Linnaeus)*	These new tools perform better than linear approaches, in particular KNN performs better than compared approaches.
Chen et al. [91]	Dengue, Malaria	LASSO	Access how the LASSO method may be useful in providing useful forecasts for different pathogens in countries with different climates.	Short-term predictions generally perform better than longer-term predictions, suggesting public health agencies may need the capacity to respond at short-notice to early warnings.
Chekol and Hagras [92]	Malaria	ANFIS and SVR	Comparing ANFIS and SVR approaches on the task of predict the malaria epidemic, in Ethiopia, up to three-months-ahead.	Similar results were observed for both approaches and they can be used in parallel.
Poirier et al. [93]	Influenza	RF, SVR and ENET	Compared internet and electronic health records data and by statistical models to identify the best approach for influenza estimates in real time.	For national and Brittany region influenza incidence rate, SVR model is the best approach.
Liang et al. [94]	Influenza	SVR	Explore the application of the SVR model for forecasting influenza data in Liaoning, China.	It is observed the feasibility of using internet search engine query data and the efficiency of SVR model in tracking the influenza.
Mollalo et al. [95]	Tuberculosis	MLP and LR	Investigated the applicability of MLP to predict the tuberculosis incidence.	Among the developed models, single hidden layer MLP had the best test accuracy.
Soliman et al. [17]	Influenza	DL, LASSO, MARS and ARIMA	Investigate utility of DL to forecast influenza in Dallas County, USA.	DL and multi-model ensemble of forecasts yield a similar competitive performance.
Chen et al. [14]	Influenza	LASSO-GP, LR, ANNs, SVR and SARIMA	Developing a non-parametric model based on LASSO-GP regression for influenza prediction considering meteorological effect.	Proposed approach outperform compared models.
Guo et al. [32]	Dengue	Ensemble	Develop an ensemble penalized regression algorithm for initializing near-real time forecasts of the dengue epidemic trajectory.	The proposed algorithm had the best performance in comparison with regression models using single penalties.
Shirmohammadi et al. [96]	Brucelossis	RF, SVR and MARS	Investigate and compared the performance of three data mining techniques to predict monthly brucellosis.	Results indicated that the RF model outperformed the SVM and MARS and it can be utilized to diagnose the behavior of brucellosis over time.
Venna et al. [97]	Influenza	ARIMA, EAKF, LSTM	Propose a novel data-driven machine learning method using long short-term memory-based multi-stage for influenza forecasting.	Proposed method performs better than the existing well-known influenza forecasting methods (ARIMA and EAKF) and the results offer a promising direction to improve influenza forecasting.
Thakur and Dharavath [98]	Malaria	ANNs	Determine the malaria abundances using clinical and environmental variables with Big Data by ANNs.	The results vary from area to area based on clinical variables and rainfall in the prediction model corresponding to areas.
Chakraborty et al. [99]	Dengue	NNAR, ARIMA, LSTM and SVR	Proposed a hybrid ARIMA-NNAR for dengue forecasting on three regions.	The ARIMA-NNAR give better forecasting accuracy in comparison to the state-of-the-art.
Su et al. [100]	Influenza	SARIMA, LASSO, LSTM and XGBoost	Develop a self-adaptive model by integrating the SARIMA and XGBoost approaches for real-time estimation of influenza in Chongqing, China.	Compared with LASSO and LSTM approaches, the proposed framework reach better improvement than this approaches for the adopted task.

2. The combination of EEMD, HTE, and WI strategy for result aggregation with weights select by NSGA-II TOPSIS allowing the enhancement of the model's performance for the adopted time series;

3. The use of HTE models for components allowing to improve the framework accuracy in relation to the use of a homogeneous ensemble of components;

Table B.1
Control Hyperparameters obtained by GS during cross-validation process for each adopted model.

State	Component	BRNN	CUBIST		GBM		QRF	PLS
		Neurons	Committes	Instances	Boosting Iterations	Max Tree Depth	Randomly Selected Predictors	Components
MG	IMF_1	(2,3,4)	(1,1,1)	(5,5,5)	(50,150,50)	(2,2,2)	(4,4,4)	(2,2,2)
	IMF_2	(1,1,1)	(1,1 ,1)	(1,0,0)	(150,150,150)	(2,3,2)	(3,3,3)	(3,3,2)
	IMF_3	(3,3,3)	(20,20,20)	(9,9,9)	(50, 150,50)	(2,1,2)	(3,3,3)	(3,3,3)
	IMF_4	(3,2,3)	(10,10,10)	(1,0,0)	(150,150,150)	(3,2,3)	(4,4,4)	(3,3,3)
	Residual	(2,1,2)	(10,10,10)	(1,0,0)	(150,50,150)	(3,2,3)	(4,3,4)	(3,3,2)
	Non-decomposed	(1,1,1)	(10,10,10)	(9,9,9)	(50, 50,50)	(1,3,2)	(4,3,2)	(2,2,2)
SP	IMF_1	(1,4,4)	(20,20,10)	(9,9,9)	(150,50,50)	(3,3,3)	(4,4,4)	(2,2,2)
	IMF_2	(1,4,4)	(1,1,1)	(0,0,0)	(150,150,150)	(2,2,2)	(4,4,4)	(3,3,3)
	IMF_3	(1,4,4)	(20,20,20)	(0,0,0)	(100,50,50)	(1,1,1)	(4,4,4)	(3,3,3)
	IMF_4	(1,4,4)	(1,1,1)	(0,0,0)	(150,150,150)	(2,3,3)	(4,4,4)	(3,3,3)
	Residual	(1,2,2)	(20,20,20)	(5,5,5)	(150,150,150)	(3,3,3)	(2,2,2)	(3,3,3)
	Non-decomposed	(3,4,4)	(10,10,10)	(5,5,5)	(150,100,100)	(3,2,3)	(4,4,4)	(3,3,3)
PR	IMF_1	(2,3,4)	(10,10,10)	(0,0,0)	(150,50,50)	(2,1,2)	(4,3,4)	(1,1,1)
	IMF_2	(1,4,4)	(1,1,1)	(0,0,0)	(150,150,150)	(3,2,3)	(4,4,4)	(3,3,3)
	IMF_3	(1,2,3)	(1,1,1)	(5,5,5)	(100,100,100)	(2,2,2)	(3,2,3)	(3,3,3)
	IMF_4	(2,2,2)	(20,20,20)	(0,0,0)	(150,150,150)	(3,2,3)	(2,2,2)	(3,3,3)
	Residual	(1,4,4)	(1,1,1)	(9,9,9)	(150,150,150)	(3,3,3)	(4,4,4)	(3,3,3)
	Non-decomposed	(1,2,3)	(1,5,5)	(5,5,5)	(50,50,50)	(2,2,3)	(4,2,3)	(3,3,3)
RJ	IMF_1	(3,3,4)	(20,20,10)	(9,9,9)	(50,100,50)	(1,1,2)	(3,3,4)	(2,2,1)
	IMF_2	(3,4,4)	(10,10,1)	(0,0,0)	(150,150,150)	(2,3,3)	(3,4,4)	(2,2,3)
	IMF_3	(3,4,3)	(1,1,1)	(5,5,5)	(150,150,100)	(3,2,2)	(3,4,3)	(3,3,3)
	IMF_4	(4,3,2)	(10,10,20)	(0,0,0)	(150,150,150)	(3,2,3)	(4,3,2)	(3,3,3)
	Residual	(4,4,4)	(20,20,1)	(5,5,9)	(100,150,150)	(2,1,3)	(4,4,4)	(3,3,3)
	Non-decomposed	(2,2,3)	(10,10,1)	(0,0,5)	(50,50,50)	(1,1,3)	(2,2,2)	(1,1,3)

Table C.1
Weights obtained by MOO approach, which are assigned for each model adopted in the signal reconstruction and models used in the structure of proposed hybrid framework.

Forecast Horizon	Weight	MG						SP					
		M1	M2	M3	M4	M5	Proposed (Models)	M1	M2	M3	M4	M5	Proposed (Model)
One-month	θ_1	1.1560	1.3087	1.4370	1.1319	1.1022	1.4536 (GBM)	1.1003	1.1498	1.2433	1.0204	1.0451	1.1457 (CUBIST)
	θ_2	0.9380	0.8953	0.9507	0.8686	1.0314	0.9076 (CUBIST)	1.0084	0.9331	0.8977	1.0061	1.0878	0.9404 (CUBIST)
	θ_3	0.7798	0.7556	0.8098	0.7723	0.8772	0.8034 (BRNN)	0.9084	1.0287	1.0712	0.9596	1.0524	1.0561 (CUBIST)
	θ_4	1.0579	1.0622	1.0897	1.0485	1.1001	1.0877 (QRF)	1.0428	1.2218	1.1090	1.0314	1.3202	1.2278 (GBM)
	θ_5	0.9954	0.9968	0.9544	0.9950	1.0373	0.9917 (GBM)	0.8722	0.9745	0.8874	0.9833	0.8774	0.9804 (GBM)
Two-months	θ_1	1.2223	1.2938	1.2598	1.2065	1.1202	1.3093 (GBM)	0.9872	1.1302	1.1864	1.1550	1.0449	1.1903 (CUBIST)
	θ_2	0.9750	0.9045	0.9448	1.0174	1.0365	1.0226 (BRNN)	0.9121	0.8508	0.9004	0.9999	1.0486	0.9527 (PLS)
	θ_3	0.7539	0.7544	0.8083	0.8209	0.8863	0.7738 (BRNN)	0.9362	1.0505	1.1512	0.9756	1.0440	1.1667 (BRNN)
	θ_4	1.0166	1.0630	1.1071	1.0289	1.0808	1.1105 (CUBIST)	1.0942	1.2205	1.0800	1.0262	1.3320	1.2825 (GBM)
	θ_5	0.9881	0.9865	0.9541	0.9835	0.9909	0.9891 (BRNN)	0.9812	0.9691	0.9780	0.8914	0.9829	0.9999 (GBM)
Three-months	θ_1	1.1500	1.3704	1.3130	1.0271	1.1125	1.3541 (GBM)	0.8756	1.0992	1.0945	1.4139	0.9896	1.1008 (CUBIST)
	θ_2	0.9111	0.8950	0.9729	0.6465	0.9666	0.9625 (CUBIST)	0.9951	0.8980	0.7734	0.6323	1.1430	0.9084 (CUBIST)
	θ_3	0.7999	0.8214	0.7725	0.6758	0.8098	0.7603 (BRNN)	0.8441	0.9291	1.2280	0.9544	1.0558	0.9713 (BRNN)
	θ_4	1.0835	1.1000	1.0323	1.0361	1.0889	1.0578 (QRF)	1.0106	1.1220	1.0468	1.0358	1.2480	1.0956 (GBM)
	θ_5	1.0024	1.0020	0.9956	1.0011	1.0966	1.0179 (QRF)	0.9122	0.9681	0.8743	0.9010	0.9040	1.0354 (GBM)

Forecast Horizon	Weight	PR						RJ					
		M1	M2	M3	M4	M5	Proposed (Models)	M1	M2	M3	M4	M5	Proposed (Model)
One-month	θ_1	0.8906	1.9995	1.1349	0.7480	1.0860	1.0751 (QRF)	1.0137	1.1176	1.4646	1.0315	1.1216	1.4744 (GBM)
	θ_2	0.8028	0.6284	0.8472	0.7622	1.0030	1.0158 (QRF)	0.9331	0.9455	0.9498	0.9149	1.0566	0.9518 (GBM)
	θ_3	0.9298	0.9333	0.8244	0.9503	1.0320	1.0065 (CUBIST)	1.0328	0.9799	1.0489	1.0215	1.0272	1.0502 (PLS)
	θ_4	0.8120	0.7744	0.8983	0.7863	0.9759	0.9686 (PLS)	1.0829	1.0986	1.0613	1.0648	1.0367	1.0807 (GBM)
	θ_5	0.9741	0.9397	0.9822	0.9688	1.0023	0.9972 (BRNN)	0.9958	0.9975	0.9706	0.9970	0.9970	0.9983 (CUBIST)
Two-months	θ_1	0.8185	1.9988	0.6224	0.6241	1.0629	0.9358 (QRF)	0.8916	1.2057	1.5927	0.8198	1.1703	0.9121 (BRNN)
	θ_2	0.7554	0.5625	0.6877	0.7407	0.9314	0.8234 (CUBIST)	0.9198	0.9425	0.9489	0.8507	1.0886	0.9051 (GBM)
	θ_3	0.9652	0.8543	0.7656	0.8175	1.0728	0.9067 (CUBIST)	0.9634	0.9550	1.0031	0.9967	1.0478	0.9746 (PLS)
	θ_4	0.7683	0.6067	0.5662	0.5488	0.9881	1.1394 (GBM)	1.0953	1.0393	1.0554	1.0076	1.0569	1.1191 (CUBIST)
	θ_5	0.9694	0.7624	0.7565	0.7307	0.9789	0.9959 (GBM)	0.9971	1.0016	0.9929	0.9962	1.0018	0.9959 (BRNN)
Three-months	θ_1	0.8343	1.9998	0.8795	0.5539	1.0713	0.8521 (QRF)	1.0679	0.9724	1.6017	1.2063	1.1190	1.0491 (BRNN)
	θ_2	0.8085	0.6458	0.7049	0.6158	0.9602	0.8281 (CUBIST)	0.9606	0.9374	0.9182	0.7245	1.1178	0.7348 (PLS)
	θ_3	0.8305	0.9981	0.6917	0.8321	1.0991	0.8287 (BRNN)	1.0310	0.9963	0.9085	0.9815	1.0073	0.9590 (PLS)
	θ_4	0.6500	0.7918	0.5345	0.5267	0.9292	1.2443 (BRNN)	1.0994	1.1006	1.1181	1.0159	1.0274	1.0157 (CUBIST)
	θ_5	0.7722	0.9420	0.7122	0.6980	0.9702	0.9997 (GBM)	0.9962	0.8994	0.9628	0.9992	0.9689	1.001 (QRF)

4. The use of WI strategy to combine the components is better than the use of DI strategy when optimization is employed to finding the set of weights;

5. The proposed approach reaches better accuracy and stability (lower standard deviation of the errors) than the compared models;

6. In 89.17% of the cases, there is a statistical difference between the errors of the EEMD-HTE-MOO framework and the compared structures.

For future works, the adoption of other decomposition techniques, such as variational mode decomposition and singular spectral analysis, is intended to reconstruct the decomposed signal through the clustering of techniques, and using artificial neural networks, adopting stacking ensemble approaches after the decomposition performance, using coyote [102,103], owls [104], and falcon [105] metaheuristics to tune hyperparameters of adopted forecasting models, using other MOEA to find the weights of WI strategy, as well as make use of the assembling of MCDM techniques. Also, is desirable to compare the recursive and direct methods to perform multi-step-ahead forecasting for the proposed task.

CRediT authorship contribution statement

Matheus Henrique Dal Molin Ribeiro: Conceptualization, Methodology, Formal analysis, Validation, Writing - original draft, Writing review & editing. **Viviana Cocco Mariani:** Conceptualization, Writing - review & editing. **Leandro dos Santos Coelho:** Conceptualization, Writing - review & editing.

Acknowledgments

Appendix A. Summary of related works

Table A.1 presents the related works regarding to machine learning and general models for epidemiologic time series forecasting.

Appendix B. Hyperparameters.

Table B.1 presents the hyperparameters obtained by GS during cross-validation process, for each component obtained by EEMD as well as for models that not considers signal decomposition. Alongside, the GS is performed only for tunable parameters of each model contained in *caret* package. Between parentheses, for each algorithm are present the parameters used on one up three-months-ahead forecast, respectively. In addition, for GBM model, the parameters *shrinkage* and *minimum number of terminal node size* are set as 0.1 and 10, respectively.

Appendix C. Weights adopted for each model in the signal reconstruction.

Table C.1 presents the weights assigned for each model adopted in the signal reconstruction, in which M1 is the model EEMD-MOO-BRNN, M2 is the model EEMD-MOO-CUBIST, M3 is the model EEMD-MOO-GBM, M4 is the model EEMD-MOO-PLS and M5 is the model EEMD-MOO-QRF. Additionally, after the weight assigned for each component of proposed model, the models employed for each component are presented.

References

[1] W.F. Wright, C.N. Pinto, K. Palisoc, S. Baghli, Viral (*aseptic*) meningitis: A review, J. Neurol. Sci. 398 (2019) 176–183, http://dx.doi.org/10.1016/j.jns.2019.01.050.

[2] A. Kwarteng, J. Amuasi, A. Annan, S. Ahuno, D. Opare, M. Nagel, et al., Current meningitis outbreak in Ghana: Historical perspectives and the importance of diagnostics, Acta Trop. 169 (2017) 51–56, http://dx.doi.org/10.1016/j.actatropica.2017.01.014.

[3] Brasil. Ministério da Saúde Departamento de Informática do Sistema Único de Saúde (DATASUS), 2019, , (Accessed 22 July 2019) (in Portuguese).

[4] H.W. Hethcote, The mathematics of infectious diseases, SIAM Rev. 42 (4) (2000) 599–653, http://dx.doi.org/10.2307/2653135.

[5] P.C. Silva, P.V. Batista, H.S. Lima, M.A. Alves, F.G. Guimarães, R.C. Silva, COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions, Chaos Solitons Fractals 139 (110088) (2020) http://dx.doi.org/10.1016/j.chaos.2020.110088.

[6] K. Sarkar, S. Khajanchi, J.J. Nieto, Modeling and forecasting the COVID-19 pandemic in India, Chaos Solitons Fractals 139 (110049) (2020) http://dx.doi.org/10.1016/j.chaos.2020.110049.

[7] B. Seidu, C. Bornaa, O.D. Makinde, An Ebola model with hyper-susceptibility, Chaos Solitons Fractals 138 (109938) (2020) http://dx.doi.org/10.1016/j.chaos.2020.109938.

[8] J.M. Aronis, N.E. Millett, M.M. Wagner, F. Tsui, Y. Ye, J.P. Ferraro, et al., A Bayesian system to detect and characterize overlapping outbreaks, J. Biomed. Inform. 73 (2017) 171–181, http://dx.doi.org/10.1016/j.jbi.2017.08.003.

[9] B. Tsanou, J.C. Kamgang, J.M.S. Lubuma, D.E.H. Danga, Modeling pyrethroids repellency and its role on the bifurcation analysis for a bed net malaria model, Chaos Solitons Fractals 136 (109809) (2020) http://dx.doi.org/10.1016/j.chaos.2020.109809.

[10] L. Vanneschi, D.M. Horn, M. Castelli, A. Popovič, An artificial intelligence system for predicting customer default in e-commerce, Expert Syst. Appl. 104 (2018) 1–21, http://dx.doi.org/10.1016/j.eswa.2018.03.025.

[11] S. Mao, B. Wang, Y. Tang, F. Qian, Opportunities and challenges of artificial intelligence for green manufacturing in the process industry, Engineering (2019) http://dx.doi.org/10.1016/j.eng.2019.08.013.

[12] S.J. Lee, Z. Xu, T. Li, Y. Yang, A novel bagging C4.5 algorithm based on wrapper feature selection for supporting wise clinical decision making, J. Biomed. Inform. 78 (2018) 144–155, http://dx.doi.org/10.1016/j.jbi.2017.11.005.

[13] L. Huang, A.L. Shea, H. Qian, A. Masurkar, H. Deng, D. Liu, Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records, J. Biomed. Inform. 99 (2019) http://dx.doi.org/10.1016/j.jbi.2019.103291.

[14] S. Chen, J. Xu, Y. Wu, X. Wang, S. Fang, J. Cheng, et al., Predicting temporal propagation of seasonal influenza using improved Gaussian process model, J. Biomed. Inform. 93 (2019) 103–144, http://dx.doi.org/10.1016/j.jbi.2019.103144.

[15] I. Savin, K. Ershova, N. Kurdyumova, O. Ershova, O. Khomenko, G. Danilov, et al., Healthcare-associated *ventriculitis* and meningitis in a neuro-ICU: Incidence and risk factors selected by machine learning approach, J. Crit. Care 45 (2018) 95–104, http://dx.doi.org/10.1016/j.jcrc.2018.01.022.

[16] D. Jiang, M. Hao, F. Ding, J. Fu, M. Li, Mapping the transmission risk of zika virus using machine learning models, Acta Trop. 185 (2018) 391–399, http://dx.doi.org/10.1016/j.actatropica.2018.06.021.

[17] M. Soliman, V. Lyubchich, Y.R. Gel, Complementing the power of deep learning with statistical model fusion: Probabilistic forecasting of influenza in Dallas County, Texas, USA, Epidemics (2019) http://dx.doi.org/10.1016/j.epidem.2019.05.004.

[18] M.H.D.M. Ribeiro, V.H.A. Ribeiro, G. Reynoso-Meza, L.S. Coelho, Multi-objective ensemble model for short-term price forecasting in corn price time series, in: International Joint Conference on Neural Networks (IJCNN). Budapest, Hungary, 2019, pp. 1–8, http://dx.doi.org/10.1109/IJCNN.2019.8851880.

[19] V.H.A. Ribeiro, G. Reynoso-Meza, Ensemble learning by means of a multi-objective optimization design approach for dealing with imbalanced data sets, Expert Syst. Appl. 147 (113232) (2020) http://dx.doi.org/10.1016/j.eswa.2020.113232.

[20] T.G. Dietterich, Ensemble methods in machine learning, in: Multiple Classifier Systems, Springer, Berlin, Germany, 2000, pp. 1–15, http://dx.doi.org/10.1007/3-540-45014-9_1.

[21] M.H.D.M. Ribeiro, L.S. Coelho, Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series, Appl. Soft Comput. 86 (105837) (2020) http://dx.doi.org/10.1016/j.asoc.2019.105837.

[22] P. Shan, Y. Zhao, Q. Wang, X. Sha, X. Lv, S. Peng, et al., Stacked ensemble extreme learning machine coupled with partial least squares-based weighting strategy for nonlinear multivariate calibration, Spectrochim. Acta A 215 (2019) 97–111, http://dx.doi.org/10.1016/j.saa.2019.02.089.

[23] M.M. Favalesso, L.M. Lorini, M.E. Peichoto, A.T.B. Guimarães, Potential distribution and ecological conditions of lonomia obliqua walker 1855 (*Saturniidae: Hemileucinae*) in Brazil, Acta Trop. 192 (2019) 158–164, http://dx.doi.org/10.1016/j.actatropica.2019.01.016.

[24] B. Bhasuran, G. Murugesan, S. Abdulkadhar, J. Natarajan, Stacked ensemble combined with fuzzy matching for biomedical named entity recognition of diseases, J. Biomed. Inform. 64 (2016) 1–9, http://dx.doi.org/10.1016/j.jbi.2016.09.009.

[25] R.G. da Silva, M.H.D.M. Ribeiro, V.C. Mariani, L.S. Coelho, Forecasting Brazilian and American covid-19 cases based on artificial intelligence coupled with climatic exogenous variables, Chaos Solitons Fractals 139 (110027) (2020) http://dx.doi.org/10.1016/j.chaos.2020.110027.

[26] Z. Wu, N.E. Huang, Ensemble empirical mode decomposition: A noise-assisted data analysis method, Adv. Adapt. Data Anal. 01 (01) (2009) 1–41, http://dx.doi.org/10.1142/S1793536909000047.

[27] J. Huan, W. Cao, Y. Qin, Prediction of dissolved oxygen in aquaculture based on EEMD and LSSVM optimized by the Bayesian evidence framework, Comput. Electron. Agricult. 150 (2018) 257–265, http://dx.doi.org/10.1016/j.compag.2018.04.022.

[28] Y.X. Wu, Q.B. Wu, J.Q. Zhu, Improved EEMD-based crude oil price forecasting using LSTM networks, Physica A 516 (2019) 114–124, http://dx.doi.org/10.1016/j.physa.2018.09.120.

[29] P. Du, J. Wang, W. Yang, T. Niu, A novel hybrid model for short-term wind power forecasting, Appl. Soft Comput. 80 (2019) 93–106, http://dx.doi.org/10.1016/j.asoc.2019.03.035.

[30] Z. Sun, Y. Chen, X. Li, X. Qin, H. Wang, A Bayesian regularized artificial neural network for adaptive optics forecasting, Opt. Commun. 382 (2017) 519–527, http://dx.doi.org/10.1016/j.optcom.2016.08.035.

[31] S.C. Worland, W.H. Farmer, J.E. Kiang, Improving predictions of hydrological low-flow indices in ungaged basins using machine learning, Environ. Model. Softw. 101 (2018) 169–182, http://dx.doi.org/10.1016/j.envsoft.2017.12.021.

[32] P. Guo, Q. Zhang, Y. Chen, J. Xiao, J. He, Y. Zhang, et al., An ensemble forecast model of dengue in Guangzhou, China using climate and social media surveillance data, Sci. Total Environ. 647 (2019) 752–762, http://dx.doi.org/10.1016/j.scitotenv.2018.08.044.

[33] A. Gusnanto, A. Ploner, F. Shuweihdi, Y. Pawitan, Partial least squares and logistic regression random-effects estimates for gene selection in supervised classification of gene expression data, J, Biomed. Inform. 46 (4) (2013) 697–709, http://dx.doi.org/10.1016/j.jbi.2013.05.008.

[34] N. Khan, S. Shahid, L. Juneng, K. Ahmed, T. Ismail, N. Nawaz, Prediction of heat waves in Pakistan using quantile regression forests, Atmos. Res. 221 (2019) 1–11, http://dx.doi.org/10.1016/j.atmosres.2019.01.024.

[35] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Trans. Evol. Comput. 6 (2) (2002) 182–197, http://dx.doi.org/10.1109/4235.996017.

[36] C.L. Hwang, K. Yoon, Multiple Attribute Decision Making, Methods and Applications, Springer, New York, 1981.

[37] P. Du, J. Wang, Z. Guo, W. Yang, Research and application of a novel hybrid forecasting system based on multi-objective optimization for wind speed forecasting, Energy Convers. Manage. 150 (2017) 90–107, http://dx.doi.org/10.1016/j.enconman.2017.07.065.

[38] X. Zhang, J. Wang, A novel decomposition-ensemble model for forecasting short-term load-time series with multiple seasonal patterns, Appl. Soft Comput. 65 (2018) 478–494, http://dx.doi.org/10.1016/j.asoc.2018.01.017.

[39] P.A. Morettin, C. Toloi, Análise de Séries Temporais, second ed., Blucher, Brazil, 2006, (in Portuguese).

[40] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, et al., The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. Lond. Ser A Math. Phys. Eng. Sci. 454 (1971) (1998) 903–995, http://dx.doi.org/10.1098/rspa.1998.0193.

[41] Z. Qian, Y. Pei, H. Zareipour, N. Chen, A review and discussion of decomposition-based hybrid models for wind energy forecasting applications, Appl. Energy 235 (2019) 939–953, http://dx.doi.org/10.1016/j.apenergy.2018.10.080.

[42] S. Wang, N. Zhang, L. Wu, Y. Wang, Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method, Renew. Energy 94 (2016) 629–636, http://dx.doi.org/10.1016/j.renene.2016.03.103.

[43] P. Jiang, Z. Liu, Variable weights combined model based on multi-objective optimization for short-term wind speed forecasting, Appl. Soft Comput. 82 (2019) 105587, http://dx.doi.org/10.1016/j.asoc.2019.105587.

[44] W.c. Wang, K.w. Chau, D.m. Xu, X.Y. Chen, Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition, Water Resour. Manage. 29 (8) (2015) 2655–2675, http://dx.doi.org/10.1007/s11269-015-0962-0.

[45] C. Wu, J. Wang, X. Chen, P. Du, W. Yang, A novel hybrid system based on multi-objective optimization for wind speed forecasting, Renew. Energy 146 (2020) 149–165, http://dx.doi.org/10.1016/j.renene.2019.04.157.

[46] D.J.C. MacKay, Bayesian Interpolation, Neural Comput. 4 (3) (1992) 415–447, http://dx.doi.org/10.1162/neco.1992.4.3.415.

[47] J.R. Quinlan, Combining instance-based and model-based learning, in: Proceedings of the 10th International Conference on International Conference on Machine Learning, ICML'93, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, ISBN: 1-55860-307-7, 1993, pp. 236–243, URL: http://dl.acm.org/citation.cfm?id=3091529.3091560.

[48] D.W. Aha, D. Kibler, M.K. Albert, Instance-based learning algorithms, Mach. Learn. 6 (1) (1991) 37–66, http://dx.doi.org/10.1007/BF00153759.

[49] J.H. Friedman, Greedy function approximation: A gradient boosting machine, Ann. Statist. 29 (5) (2001) 1189–1232.

[50] N. Meinshausen, Quantile regression forests, J. Mach. Learn. Res. 7 (Jun) (2006) 983–999, URL: https://jmlr.org/papers/volume7/meinshausen06a/meinshausen06a.pdf, (Accessed 20 August 2019).

[51] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, http://dx.doi.org/10.1023/A:1010933404324.

[52] K. Vaysse, P. Lagacherie, Using quantile regression forest to estimate uncertainty of digital soil mapping products, Geoderma 291 (2017) 55–64, http://dx.doi.org/10.1016/j.geoderma.2016.12.017.

[53] S. Wold, H. Martens, H. Wold, The multivariate calibration problem in chemistry solved by the pls method, in: B. Kågström, A. Ruhe (Eds.), Matrix Pencils, Springer, Berlin, Germany, 1983, pp. 286–293, http://dx.doi.org/10.1007/BFb0062108.

[54] P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial, Anal. Chim. Acta 185 (1986) 1–17, http://dx.doi.org/10.1016/0003-2670(86)80028-9.

[55] R. Marler, J. Arora, Survey of multi-objective optimization methods for engineering, Struct. Multidiscip. Optim. 26 (6) (2004) 369–395, http://dx.doi.org/10.1007/s00158-003-0368-6.

[56] G. Reynoso-Meza, X. Blasco, J. Sanchis, M. Martínez, Controller tuning using evolutionary multi-objective optimisation: Current trends and applications, Control Eng. Pract. 28 (2014) 58–73, http://dx.doi.org/10.1016/j.conengprac.2014.03.003.

[57] J. Neville, D. Jensen, Bias/variance analysis for relational domains, in: H. Blockeel, J. Ramon, J. Shavlik, P. Tadepalli (Eds.), Inductive Logic Programming, Springer, Berlin, Germany, 2008, pp. 27–28, http://dx.doi.org/10.1007/978-3-540-78469-2_6.

[58] S. Mirjalili, Genetic algorithm, in: Evolutionary Algorithms and Neural Networks: Theory and Applications, Springer, Cham, 2019, pp. 43–55, http://dx.doi.org/10.1007/978-3-319-93025-1_4.

[59] T.C. Bora, V.C. Mariani, L.S. Coelho, Multi-objective optimization of the environmental-economic dispatch with reinforcement learning based on non-dominated sorting genetic algorithm, Appl. Therm. Eng. 146 (2019) 688–700, http://dx.doi.org/10.1016/j.applthermaleng.2018.10.020.

[60] V.H.A. Ribeiro, Reynoso-Meza. G., A holistic multi-objective optimization design procedure for ensemble member generation and selection, Appl. Soft Comput. 83 (105664) (2019) http://dx.doi.org/10.1016/j.asoc.2019.105664.

[61] J.X.V. Neto, E.J.G. Junior, S.R. Moreno, H.V.H. Ayala, V.C. Mariani, L.S. Coelho, Wind turbine blade geometry design based on multi-objective optimization using metaheuristics, Energy 162 (2018) 645–658, http://dx.doi.org/10.1016/j.energy.2018.07.186.

[62] G.H. Tzeng, J.J. Huang, Multiple Attribute Decision Making: Methods and Applications, Chapman and Hall/CRC, London, United Kingdom, 2011.

[63] C.L. Hwang, Y.J. Lai, T.Y. Liu, A new approach for multiple objective decision making, Comput. Oper. Res. 20 (8) (1993) 889–899, http://dx.doi.org/10.1016/0305-0548(93)90109-V.

[64] F.M. Pouzols, A.B. Barros, Automatic clustering-based identification of autoregressive fuzzy inference models for time series, Neurocomputing 73 (10) (2010) 1937–1949, http://dx.doi.org/10.1016/j.neucom.2009.11.028.

[65] A. Veloz, R. Salas, H. Allende-Cid, H. Allende, C. Moraga, Identification of lags in nonlinear autoregressive time series using a flexible fuzzy model, Neural Process. Lett. 43 (3) (2016) 641–666, http://dx.doi.org/10.1007/s11063-015-9438-1.

[66] S. Ma, R. Fildes, Forecasting third-party mobile payments with implications for customer flow prediction, Int. J. Forecast. 36 (3) (2020) 739–760, http://dx.doi.org/10.1016/j.ijforecast.2019.08.012.

[67] M. Rana, A. Rahman, Multiple steps ahead solar photovoltaic power forecasting based on univariate machine learning models and data re-sampling, Sust. Energy Grids Netw. 21 (100286) (2020) http://dx.doi.org/10.1016/j.segan.2019.100286.

[68] P. Xue, Y. Jiang, Z. Zhou, X. Chen, X. Fang, J. Liu, Multi-step ahead forecasting of heat load in district heating systems using machine learning algorithms, Energy 188 (116085) (2019) http://dx.doi.org/10.1016/j.energy.2019.116085.

[69] R.J. Hyndman, G. Athanasopoulos, Forecasting: principles and practice, in: OTexts, second ed., ISBN: 9780987507112, 2018, URL https://otexts.org/fpp2/, (Accessed 06 August 2019).

[70] J. Song, J. Wang, H. Lu, A novel combined model based on advanced optimization algorithm for short-term wind speed forecasting, Appl. Energy 215 (2018) 643–658, http://dx.doi.org/10.1016/j.apenergy.2018.02.070.

[71] F.X. Diebold, R.S. Mariano, Comparing predictive accuracy, J. Bus. Econom. Statist. 13 (3) (1995) 253–263, http://dx.doi.org/10.1080/07350015.1995.10524599.

[72] R Core Team, R: A Language and Environment for Bus. Comput, R Foundation for Statistical Computation;, Vienna, Austria, 2018, URL http://www.R-project.org/.

[73] D.C. Bowman, J.M. Lees, The Hilbert-Huang transform: A high resolution spectral method for nonlinear and nonstationary time series, Seismol. Res. Lett. 84 (6) (2013) 1074–1080, http://dx.doi.org/10.1785/0220130025.

[74] M. Kuhn, Building predictive models in r using the caret package, J. Stat. Softw. Artic 28 (5) (2008) 1–26, http://dx.doi.org/10.18637/jss.v028.i05.

[75] O. Mersmann, mco: Multiple Criteria Optimization Algorithms and Related Functions, R Package Version 1.0-15.1, 2014, URL https://CRAN.R-project.org/package=mco.

[76] B.A.C. Martin, MCDM: Multi-Criteria Decision Making Methods for Crisp Data, R Package Version 1.2, 2016, URL https://CRAN.R-project.org/package=MCDM.

[77] S. Zhang, J. Wang, Z. Guo, Research on combined model based on multi-objective optimization and application in time series forecast, Soft Comput. (2018) http://dx.doi.org/10.1007/s00500-018-03690-w.

[78] Y. Liu, X. Yao, Ensemble learning via negative correlation, Neural Netw. 12 (10) (1999) 1399–1404, http://dx.doi.org/10.1016/S0893-6080(99)00073-8.

[79] J. Heinermann, O. Kramer, Machine learning ensembles for wind power prediction, Renew. Energy 89 (2016) 671–679, http://dx.doi.org/10.1016/j.renene.2015.11.073.

[80] Y. Xiang, L. Gou, L. He, S. Xia, W. Wang, A SVR–ANN combined model based on ensemble EMD for rainfall prediction, Appl. Soft Comput. 73 (2018) 874–883, http://dx.doi.org/10.1016/j.asoc.2018.09.018.

[81] M. Santhosh, C. Venkaiah, D.V. Kumar, Short-term wind speed forecasting approach using ensemble empirical mode decomposition and deep Boltzmann machine, Sustain. Energy Grids Netw. 19 (2019) 100242, http://dx.doi.org/10.1016/j.segan.2019.100242.

[82] S.F. Stefenon, M.H.D.M. Ribeiro, A. Nied, V.C. Mariani, L.S. Coelho, D.F.M. da Rocha, et al., Wavelet group method of data handling for fault prediction in electrical power insulators, Int. J. Electr. Power Energy Syst. 123 (2020) 106269, http://dx.doi.org/10.1016/j.ijepes.2020.106269.

[83] Z. Qu, K. Zhang, W. Mao, J. Wang, C. Liu, W. Zhang, Research and application of ensemble forecasting based on a novel multi-objective optimization algorithm for wind-speed forecasting, Energy Convers. Manage. 154 (2017) 440–454, http://dx.doi.org/10.1016/j.enconman.2017.10.099.

[84] L.T. Bui, V.T. Vu, T.T.H. Dinh, A novel evolutionary multi-objective ensemble learning approach for forecasting currency exchange rates, Data Knowl. Eng. 114 (2018) 40–66, http://dx.doi.org/10.1016/j.datak.2017.07.001.

[85] S.B. Taieb, G. Bontempi, A.F. Atiya, A. Sorjamaa, A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition, Expert Syst. Appl. 39 (8) (2012) 7067–7083, http://dx.doi.org/10.1016/j.eswa.2012.01.039.

[86] T. Xiong, C. Li, Y. Bao, Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: Evidence from the vegetable market in China, Neurocomputing 275 (2018) 2831–2844, http://dx.doi.org/10.1016/j.neucom.2017.11.053.

[87] S.R. Moreno, R.G. da Silva, V.C. Mariani, L.S. Coelho, Multi-step wind speed forecasting based on hybrid multi-stage decomposition model and long short-term memory neural network, Energy Convers. Manage. 213 (112869) (2020) http://dx.doi.org/10.1016/j.enconman.2020.112869.

[88] S. Ch, S. Sohani, D. Kumar, A. Malik, B. Chahar, A. Nema, et al., A support vector machine-firefly algorithm based forecasting model to determine malaria transmission, Neurocomputing 129 (2014) 279–288, http://dx.doi.org/10.1016/j.neucom.2013.09.030.

[89] P. Guo, T. Liu, Q. Zhang, L. Wang, J. Xiao, Q. Zhang, et al., Developing a dengue forecast model using machine learning: A case study in China, PLoS Negl. Trop. Dis. 11 (10) (2017) 1–22, http://dx.doi.org/10.1371/journal.pntd.0005973.

[90] J.M. Scavuzzo, F. Trucco, M. Espinosa, C.B. Tauro, M. Abril, C.M. Scavuzzo, et al., Modeling dengue vector population using remotely sensed data and machine learning, Acta Trop. 185 (2018) 167–175, http://dx.doi.org/10.1016/j.actatropica.2018.05.003.

[91] Y. Chen, C.W. Chu, M.I. Chen, A.R. Cook, The utility of LASSO-based models for real time forecasts of endemic infectious diseases: A cross country comparison, J. Biomed. Inform. 81 (2018) 16–30, http://dx.doi.org/10.1016/j.jbi.2018.02.014.

[92] B.E. Chekol, H. Hagras, Employing machine learning techniques for the malaria epidemic prediction in ethiopia, in: Computer Science and Electronic Engineering (CEEC). Colchester, United Kingdom, 2018, pp. 89–94, http://dx.doi.org/10.1109/CEEC.2018.8674210.

[93] C. Poirier, A. Lavenu, V. Bertaud, B. Campillo-Gimenez, E. Chazard, M. Cuggia, et al., Real time influenza monitoring using hospital big data in combination with machine learning methods: Comparison study, JMIR Publ. Health Surveill. 4 (4) (2018) e11361, http://dx.doi.org/10.2196/11361.

[94] F. Liang, P. Guan, W. Wu, D. Huang, Forecasting influenza epidemics by integrating internet search queries and traditional surveillance data with the support vector machine regression model in liaoning, from 2011 to 2015, PeerJ 6 (2018) http://dx.doi.org/10.7717/peerj.5134.

[95] A. Mollalo, L. Mao, P. Rashidi, G.E. Glass, A GIS-based artificial neural network model for spatial distribution of tuberculosis across the continental United States, Int. J. Environ. Res. Publ. Health 16 (1) (2019) http://dx.doi.org/10.3390/ijerph16010157.

[96] K.N. Shirmohammadi, L. Tapak, O. Hamidi, Z. Maryanaji, A comparison of three data mining time series models in prediction of monthly brucellosis surveillance data, Zoonoses Publ. Health (2019) 1–14, http://dx.doi.org/10.1111/zph.12622.

[97] S.R. Venna, A. Tavanaei, R.N. Gottumukkala, V.V. Raghavan, A.S. Maida, S. Nichols, A novel data-driven model for real-time influenza forecasting, IEEE Access 7 (2019) 7691–7701, http://dx.doi.org/10.1109/ACCESS.2018.2888585.

[98] S. Thakur, R. Dharavath, Artificial neural network based prediction of malaria abundances using big data: A knowledge capturing approach, Clin. Epidemiol. Glob. Health 7 (1) (2019) 121–126, http://dx.doi.org/10.1016/j.cegh.2018.03.001.

[99] T. Chakraborty, S. Chattopadhyay, I. Ghosh, Forecasting dengue epidemics using a hybrid methodology, Physica A 527 (2019) 121266, http://dx.doi.org/10.1016/j.physa.2019.121266.

[100] K. Su, L. Xu, G. Li, X. Ruan, X. Li, P. Deng, et al., Forecasting influenza activity using self-adaptive ai model and multi-source data in Chongqing, China, E Bio. Med. (2019) http://dx.doi.org/10.1016/j.ebiom.2019.08.024.

[101] M.F. Li, X.P. Tang, W. Wu, H.B. Liu, General models for estimating daily global solar radiation for different solar radiation zones in mainland China, Energy Convers. Manage. 70 (2013) 139–148, http://dx.doi.org/10.1016/j.enconman.2013.03.004.

[102] J. Pierezan, L.S. Coelho, Coyote optimization algorithm: A new metaheuristic for global optimization problems, in: IEEE Congress on Evolutionary Computation (CEC). Rio de Janeiro, Brazil, 2018, pp. 2633–2640, http://dx.doi.org/10.1109/CEC.2018.8477769.

[103] J. Pierezan, G. Maidl, E.M. Yamao, L.S. Coelho, V.C. Mariani, Cultural coyote optimization algorithm applied to a heavy duty gas turbine operation, Energy Convers. Manage. 199 (111932) (2019) http://dx.doi.org/10.1016/j.enconman.2019.111932.

[104] E.H. de Vasconcelos Segundo, V.C. Mariani, L.S. Coelho, Metaheuristic inspired on owls behavior applied to heat exchangers design, Therm. Sci. Eng. Prog. 14 (100431) (2019) http://dx.doi.org/10.1016/j.tsep.2019.100431.

[105] E.H. de Vasconcelos Segundo, V.C. Mariani, L.S. Coelho, Design of heat exchangers using falcon optimization algorithm, Appl. Therm. Eng. 156 (2019) 119–144, http://dx.doi.org/10.1016/j.applthermaleng.2019.04.038.