Forecasting epidemiological time series based on decomposition and optimization approaches

Matheus H. D. M. Ribeiro*† ,Ramon G. Silva*, Naylene Fraccanabbia‡, Viviana C. Mariani‡§ and Leandro S. Coelho*§
*Industrial and Systems Engineering Graduate Program (PPGEPS), Pontifical Catholic University of Parana (PUCPR),
Curitiba, Parana, Brazil. 80215-901
†Department of Mathematics, Federal Technological University of Parana (UTFPR),
Pato Branco, Parana, Brazil. 85503-390
‡Mechanical Engineering Graduate Program (PPGEM), Pontifical Catholic University of Parana (PUCPR),
Curitiba, Parana, Brazil. 80215-901
§Department of Electrical Engineering, Federal University of Parana (UFPR),
Curitiba, Parana, Brazil. 81530-000

Emails: matheus.dalmolinribeiro@gmail.com, gomes.ramon@pucpr.edu.br, nayfraccanabbia@outlook.com,
viviana.mariani@pucpr.br, leandro.coelho@pucpr.br

***Abstract*—Epidemiological time series forecasting plays an important role in health public system, since it allows managers to develop strategic planning to avoid possible epidemics. In this aspect, a hybrid approach is developed to forecast confirmed cases of megingitis in the Para, Parana and Santa Catarina states, Brazil. In this case, ensemble empirical mode decomposition (EEMD) is applied to decompose the original signal, quantile random forests (QRF) is adopted to forecast each component obtained in decomposition stage and multi-objective optimization (MOO) is used to reconstruct the final forecasting. To assess the performance of adopted methodology, comparisons are conducted with approach that considers to reconstruct the signal by simple sum (EEMD-QRF) and QRF without decomposition. In this context criteria such as mean squared error, symmetric mean absolute percentage error and coefficient of determination as well as statistical tests are adopted. As results, EEMD-QRF-MOO reached lower errors and better coefficient of determination in most of the cases. Indeed, the EEMD-QRF-MOO and EEMD-QRF squared errors are statistical equals, and lower than QRF squared errors. With these results it is conclude that using decomposition technique combined with machine learning models and optimization approach can be adopted to enhance the model performance, whose results may be used to perform accurate forecasting.**

***Keywords*—Decomposition, ensemble, time series, meningitis, multi-objective optimization.**

I. Introduction

Meningitis is an inflammation that has several classifications with specific causes and symptoms, and unfortunately it is still a major public health problem, causing irreversible health damage with high mortality rates [1]. Prevention, early diagnosis and initiation of treatment are fundamental to the good prognosis of the disease. In addition, in the field of diseases which are public health problems some studies have already been conducted for regression and classification tasks using machine learning approaches [2]–[5].

Time series forecasting, a subfield of regression area, aims to using past data to forecast future values with purpose of, for example, making a strategic planning to improve the knowledge in the domain that are inserted and developed public policies. Developing an efficient model is desirable and techniques such as ensemble and decomposition can be used for this purpose. These strategies can be employed to deal with nonlinearity, nonstationarity and cyclicity inherent to time series.

With relation of methodologies adopted to enhance models performance, ensemble approach can be used for this purpose and is applicable for regression [6] and classification [7] tasks. The main aspect of this approach lies on training several base (weak) models and combine its predictions to build an efficient model [8]. It is supposed that this improvement occurs because each base model learns different characteristics of the data and add these information in the final results. Indeed, this methodology has proven effective in forecasting tasks in different knowledge domains [9]–[11].

An additional approach, usually adopted to improve the models performance, is the ensemble empirical mode decomposition (EEMD). Employing decomposition the original signal is separated into components and extract relevant information from them. Decomposition based on EEMD methodology consists of sifting an ensemble of white noise-added signal and treats the mean as the final true result. Components obtained are namely intrinsic mode functions (IMF) and residual [12]. In this case, each IMF is treated as an input set and are trained separated using some algorithm. After this, it is necessary reconstruct the original signal. Besides, the EEMD approach has proven effective in forecasting task [13]–[15].

Considering the aforementioned, the aim of this paper is to employ an approach that combine ensemble (quantile random forests - QRF) and decomposition (EEMD) to forecast one month ahead the meningitis confirmed cases number in the Para (PA), Parana (PR) and Santa Catarina (SC) states, Brazil. The QRF approach is adopted, once it showed good results

in applications from different areas [16]–[18]. The QRF is an extension of random forest model, which keeps all information contained in the leaves with objective to estimate conditional probabilities. Due to the dynamic behavior of this disease, the short-term forecasting is adopted. In the most of the cases, a simple sum is adopted to reconstruct the original signal. This approach considers same weight for all components, which penalizes those that explain more the data variability by attribute the same importance for all components. Faced with this, in this paper multi-objective optimization (MOO) is adopted to reconstruct the original signal, with purpose to choose different weights to each component. The MOO approach enable to deal with bias-variance trade-off and improve the models' accuracy and stability simultaneously [19].

The contributions of this paper are twofold. The first refers to the fact that there are limited discussions regarding how to reconstruct the original signal using different approaches as presented by [20], [21] and [22]. In this case, an investigation of effectiveness of MOO to aggregate the results of EEMD components is conducted. In this context this paper is seeking to add discussions in this field. The second fold lies on the development of a hybrid forecast model, to forecast the values for a disease that over the years can affect the population and the results may be used to conduct public polices and preventive campaigns.

The remainder of this paper is structured as follows: Section II-A presents the data sets adopted in this paper. Section II-B describes the methods employed in this paper. Section III presents the data modeling steps. Section IV shows the results and discussions. Finally, Section V concludes this paper and presents the proposals of future research.

II. Material and Methods

In this section, the data as well as the adopted methods used in this paper are presented.

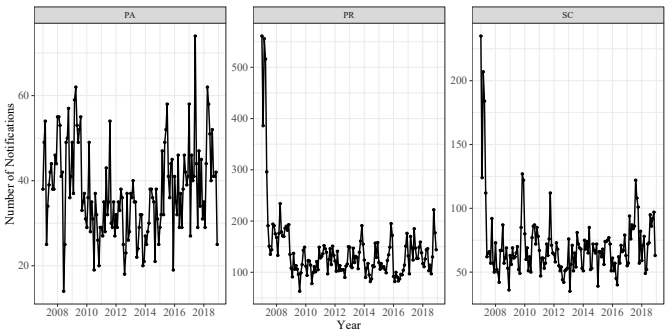
A. Material

The data used in this paper refers to monthly records of meningitis confirmed cases number, recorded in disease information system. The information for PA, PR and SC states, Brazil, are chosen because the disease behavior is different and it enable to evaluate the proposed approach in some scenarios. Table I presents the statistical indicators.

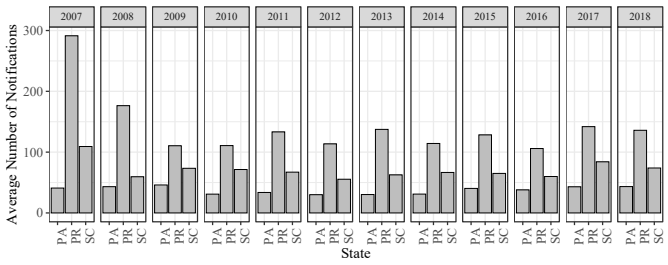
TABLE I
Statistical Measures for the Meningitis Notified Cases Number for All States.

State	Description	n	Minimum	Mean	Maximum	Standard Deviation
Para	All data	144	14	37.54	74	10.66
	Training	132	14	37.02	74	10.51
	Test	12	25	43.33	62	11.01
Parana	All data	144	63	141.63	561	71.46
	Training	132	63	142.15	561	73.95
	Test	12	97	135.92	222	35.06
Santa Catarina	All data	144	35	70.68	235	26.78
	Training	132	35	70.39	235	27.56
	Test	12	49	73.92	97	16.30

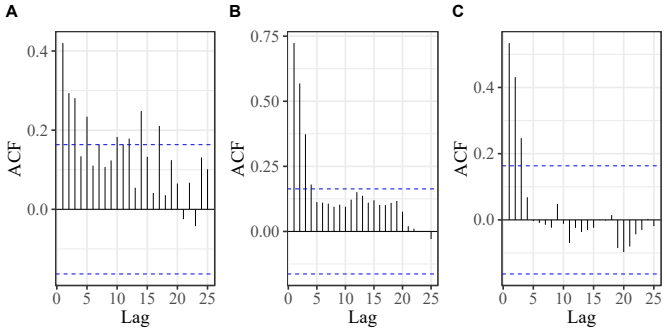
For three adopted series, information between years 2007 and 2018 are available on Department of Informatics of the Unified Health System (*Departamento de Informática do Sistema Único de Saúde*, DATASUS, in Portuguese) database [23], which are used in this paper. In this case, the first eleven years are used for model training, while the last year is used for performance testing.



(a) Observed series.



(b) Notifications number for each state over twelve years.



(c) Autocorrelations (ACF) for PA (A), PR (B) and SC (C) states

Fig. 1. Graphical analysis for adopted time series.

Figures 1a and 1b illustrate the data sets behavior, in which for PR state, there is a greater number of notifications than other states. The Augmented Dickey-Fuller test shows that the three series are non-stationary (DF = -5.35 - -3.41, p-value $>$ 0.05). With objective of evaluate the presence of seasonality in the data, Kruskal-Wallis test is performed. In this case, for PA and SC series, there is no evidence of seasonality (χ^2_{11} = 13.32 - 15.50, p-value $>$ 0.05), while for the series related with PR state, there is evidence of seasonality (χ^2_{11} = 33.07, p-value $<$ 0.05) [24]. Additionally, Figure 1c suggests that up to first four observations are correlated, and can be used as inputs for the data modeling.

B. Methods

1) Ensemble Empirical Mode Decomposition : The EEMD technique was proposed by [12] and is an extension of Empirical Mode Decomposition (EMD) algorithm. This approach consists of sifting an ensemble of white noise-added signal (data) and treats the mean as the final true result. In this sense, it is performed the decomposition of time series signal with objective to extract the coexisting oscillatory functions, named IMF and residual component, from original data. After this, the ensemble average of corresponding IMF is treated as the final decomposed time series. Further details, [20] can be consulted. It was proposed to overcome the drawback of the so-called "mode mixing" problem (MMP), which is considered an advantage of this approach. The MMP it is known as the fact that each single IMF consists of signals with dramatically disparate scales or a signal of the same scale appears in different IMF components [25]. Faced with this, two disadvantages can be stated, such as: (i) Extra noise exists in the reconstructed signal and (ii) it needs more computational resources [26].

2) Quantile Random Forests : The quantile random forests (QRF) [27] approach is an extension of random forests (RF) ensemble model [28]. It provides information about the full conditional distribution of the response variable, not only about the conditional mean. In this approach, the use of conditional quantil is to enhance the RF performance, which makes this a consistent approach [27]. The main assumption about QRF lies on that weighted observations can be used for estimating the conditional mean [16]. Additionally, while the RF approach keeps in the results information as regards the notifications number average of the leaves, the QRF keeps all notifications contained in the leaves with objective of to estimate

$$\hat{P}\left(C_{t+h} \leq C | p_t\right) = \hat{F}(C|p_t), \tag{1}$$

in which the left side of equality represents the conditional probability of notifications number in relation to predictors, being C_{t+h} the notifications in a h horizon and p_t the predictors' number, (t = $1, \ldots, N$). The right side refers to the conditional distribution function (CDF) of notifications number average regarding to predictors. Considering that QRF uses the quantiles in the predictions process, the α-quantil of CDF is stated as the probability that the number of notifications is less than Q_α given p_t is equals to α, whose estimate of α is stated as follows,

$$\hat{Q}_\alpha(p_t) = inf\left\{C : \hat{F}(C|p_t) \geq \alpha\right\}, \tag{2}$$

in which $\hat{Q}_\alpha(p_t)$ is the α-quantil estimate with respet to predictor t [27].

3) Multi-objective optimization : On some problems, it is necessary to minimize (or maximize) multiple objectives in order to achieve a preferable solution. Indeed, if there are two or more objectives, a methodology that can be used for this purpose is the MOO [29]. The MOO is performed basically on three folds: First, it is necessary to define the multi-objective problem (MOP); Second, some algorithm is 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) [30]. According to [29], in the first step, the MOP is defined, which decision variables, constraints and objectives are stated. With objective to obtain an accurate and stable model, the bias-variance framework may be adopted as objectives [31]. Hence, it is expected that the final model presents results with lower error and variance. After the aforementioned step, in the MOO step, an optimization algorithm is applied with objective to find the Pareto Front approximation (PF). This set is composed of non-dominated solutions, which are solutions for which there is no other permissible solution that simultaneously improves all the objective functions without sacrificing at least one other objective function. In this aspect, each set of decision variables associated to each element of PF makes up the Pareto Set (PS) [29]. Lastly, in the MCDM step, it is found a preferable set of decision variables that permits to deal with the trade-off between the objectives.

III. Methodology

The steps of adopted methodology are similar the procedure of [32] and summarized as follow,

1) Performing EEMD to data set and obtain 4 IMFs and residual (R) component;
2) For each component, the autocorrelation analyzes showed, in most of the cases, that up to four lags are suitable to use as predictors. Without loss of generality, for all components and states studied, this configuration is used.
3) Applying center-scale (subtracts the mean and divide by standard deviation of variable) preprocessing;
4) Training each IMF and R components using QRF using leave-one-out cross validation with time slice window, according to follow structure

$$y_{(t,k)} = f\left\{y_{(t-1,k)}, y_{(t-2,k)}, y_{(t-3,k)}, y_{(t-4,k)}\right\} + \epsilon, \tag{3}$$

in which f is a function related to adopted model for training process, $y_{(t,k)}$ is the forecast value for *k*-th component obtained in decomposition stage ($k = 1, \ldots, 5$) at time t, $y_{(t-d,k)}$ are the previous notified cases lagged in $d = 1, \ldots, 4$ and ϵ is the random error, which follows a normal distribution with zero mean and variance σ^2.

5) Reconstructing the original signal, based on training predictions, using MOO process, as follows:
   a) In the MOP, the cost function is stated as follows:

$$\hat{y}_t = \sum_{i=1}^{k=4} \theta_i \widehat{IMF}_{QRF_k} + \theta_5 \widehat{R}_{QRF_5}, \tag{4}$$

in which $\hat{y}_t$ is the predicted value at time t, $\widehat{IMF}_{QRF_k}$ and $\widehat{R}_{QRF_5}$ are the predictions of each component and θ_i is the weight to be estimated,

in which $\theta_i \sim Unif[-2, 2]$. The objectives are defined as follow,

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

where

$$E(\hat{\mathbf{y}}) = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i \quad \text{and} \quad \mathbf{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{5}$$

in which $\hat{y}_i$ and y_i are *i*-th predicted and observed values.

b) In the sequence the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) [33] is applied and the Pareto Front approximation (PF) is obtained. The crossover rate is set to 0.9 because it permits new structures to be introduced into the population at a faster rate, whereas mutation rate is set to 0.1 because it prevents a given position from becoming stagnant in a set of values for the parameters to be optimized;

c) With purpose to find the best set of weights, the technique to order preference by similarity (TOPSIS) [34] is employed, in which the weights for bias (error) and variance objectives are 10% and 90%, respectively;

6) Forecasting the meningitis confirmed cases number, according to

$$\hat{y}_t = \sum_{i=1}^{k=4} \hat{\theta}_k \widehat{IMF}_{QRF_k} + \hat{\theta}_5 \widehat{R}_{QRF_5}, \tag{6}$$

in which $\hat{\theta}_k$ are the estimated weights.

7) Computing the performance measures such as coefficient of determination (R^2), root mean square error (RMSE) and symmetric mean absolute percentage error (sMAPE), which are stated as follow,

$$R^2 = 1 - \frac{\sum_{i=1}^{n}[y_i - \hat{y}_i]^2}{\sum_{i=1}^{n}[y_i - \overline{y}_i]^2}, \tag{7}$$

$$sMAPE = 100 \times \sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{(|y_i| + |\hat{y}_i|/2)}\right|, \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{9}$$

in which n represents the number of observations of the training and test sets, y_i is the *i*-th observed value and $\hat{y}_i$ is the *i*th predicted value by the adopted model.

Friedman test can be applied to evaluate the squared errors (SE). This approach is able to show if there is difference between the average ranks for more than two samples [35]. It can be employed to answer if on a set of k algorithms (greater than two), do at least two of the methods present different results.

Second to [35], if the null hypothesis is rejected, it is necessary to apply a post-hoc test to find which groups have different results. Hence, the Nemenyi test can be applied. In this approach a threshold is obtained according to (10)

$$CD = \frac{q_{\infty,k,\alpha}}{\sqrt{2}}\sqrt{\frac{k(k+1)}{6}}, \tag{10}$$

in which CD is the critical difference for which allows to infer that the results (errors) are statistically equals or not for the compared groups, $q_{\infty,k,\alpha}$ are the quantiles based on the studentized range statistic, k is the groups number and n is the observations number. If the critical differences rank sums $|R_i - R_j|$ are greater than CD, there is difference between results from groups i and j [36]–[38].

In addition, Figure 2 presents the framework of the proposed hybrid forecasting approach.

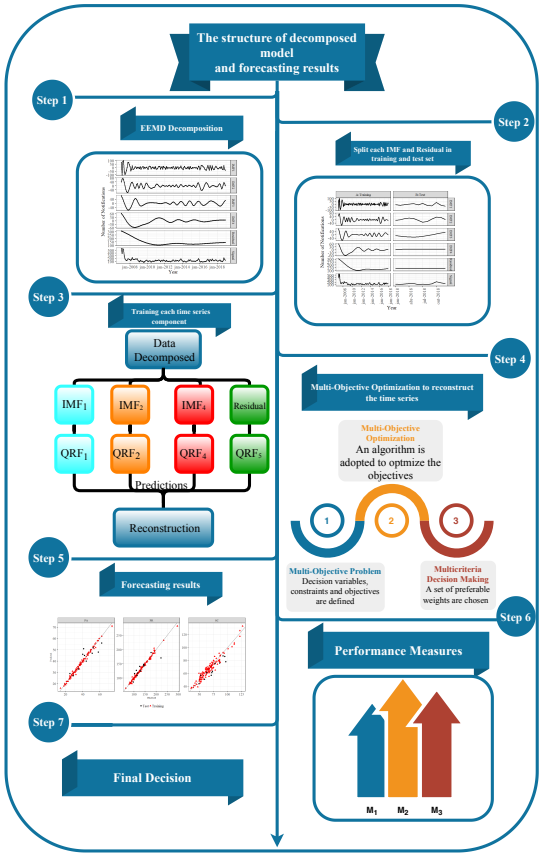
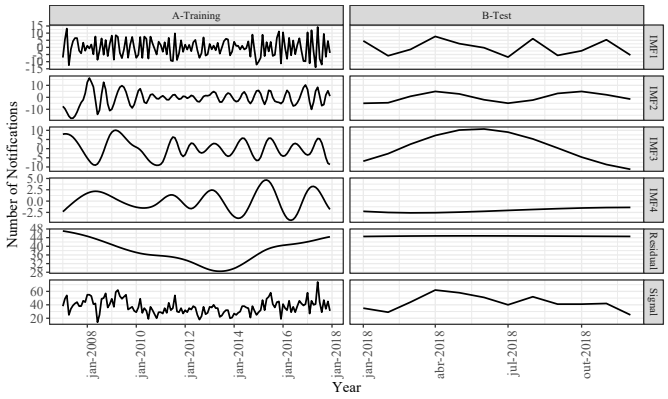


Fig. 2. Framework of the proposed hybrid forecasting approach.

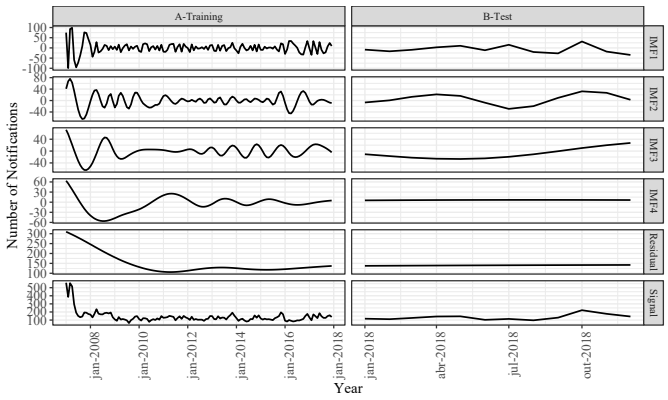
The results presented in Section IV are generated using the processor Intel(R) Core(TM) i5-4200U central processing unit of 1.6Hz in Windows 10 operating system. The R software [39] is adopted to perform the modeling.

IV. Results

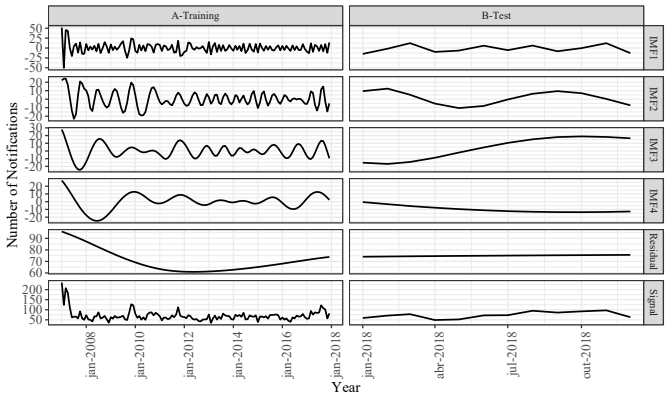
For adopted series, Figure 3 represents the decomposed series for each state. Faced with this, four IMFs and one residual component are obtained.



(a) Decomposed series for PA state.



(b) Decomposed series for PR state.



(c) Decomposed series for SC state.

Fig. 3. Decomposed time Series for each State.

Table II presents the number of randomly selected predictors (*mtry*) for each component/model and time series. On most of the cases all predictors, or lags, are used in QRF training.

TABLE II
CONTROL HYPERPARAMETERS (*mtry*) EMPLOYED IN EACH MODEL.

Component/Model	PA	PR	SC
IMF_1	2	4	3
IMF_2	4	4	4
IMF_3	4	2	3
IMF_4	4	2	4
Residual	4	3	4
QRF	4	4	2

Figure 4 illustrates the PF for each MOO conducted, as well as the selected solution (red triangle) and non-selected solution (black circle) by TOPSIS approach.

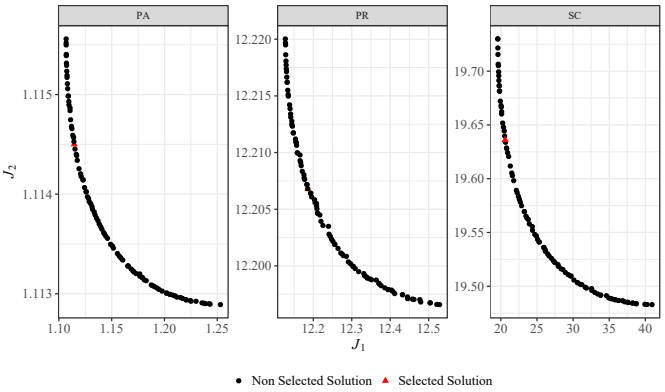


Fig. 4. PF for each EEMD-QRF-MOO for PA (left), PR (middle) and SC (right).

According Figure 4, there is trade-off for bias-variance, what justifies the use of the MOO process. For PR and SC greater variation and error are observed, with less intensity for PA state.

Table III presents the weights set obtained by TOPSIS technique after performing MOO.

TABLE III
WEIGHTS ADOPTED FOR EACH IMF AND RESIDUAL IN THE SIGNAL RECONSTRUCTION.

State	θ_1	θ_2	θ_3	θ_4	θ_5
PA	1.1020	1.0409	1.0151	1.0126	1.0146
PR	1.0904	1.0214	1.0143	0.9798	0.9938
SC	1.0691	1.0162	0.9118	1.1994	0.9892

Considering the Table III results, for PR and PA the first component received greater importance, while for SC state, the component four had more importance. This suggest that adopting same weight for all components can lead to less accurate result. Theses weights are determined on the basis of the NSGA-II algorithm, with objective to acquire better weight coefficients of those several single models.

The models' performance measures to forecast one month ahead the meningitis notified cases number for each state are shown in Table IV. Best results are stated in bold.

TABLE IV
PERFORMANCE MEASURES ADOPTED IN THE MODELS EVALUATION.

State	Measure	EEMD-QRF	EEMD-QRF-MOO	QRF
Training set				
PA	RMSE	1.3550	**1.0825**	3.6596
	R^2	0.9882	**0.9893**	0.8919
	sMAPE	2.57%	**2.08%**	4.19%
PR	RMSE	4.0243	**3.6518**	7.6555
	R^2	0.9867	**0.9901**	0.8552
	sMAPE	**1.93%**	2.03%	2.47%
SC	RMSE	4.9434	**4.4362**	7.0872
	R^2	0.9148	**0.9317**	0.8552
	sMAPE	6.50%	**5.51%**	4.04%
Test set				
PA	RMSE	6.0690	**6.0346**	12.7777
	R^2	0.6693	**0.6751**	0.0369
	sMAPE	11.73%	**11.57%**	23.54%
PR	RMSE	17.5855	**17.3518**	31.8764
	R^2	0.7323	**0.7378**	0.1869
	sMAPE	8.24%	**7.97%**	15.28%
SC	RMSE	**8.1240**	9.6220	15.4407
	R^2	0.8472	**0.8628**	0.1355
	sMAPE	**9.82%**	12.03%	19.73%

Considering the results pointed out in Table IV, the EEMD-QRF-MOO approach outperform other two approaches for both training set and test set (on two data sets). Expressive improvement is observed when results of EEMD-QRF-MOO and QRF are compared. By analyzing the performance of EEMD-QRF-MOO and EEMD-QRF, the results are similar, but best for first modeling. Evaluating the results in test set, with regard the meningitis notified cases number for PA and PR states, the approach that combine signal decomposition and optimization approaches reached an improvement for RMSE and sMAPE ranged between 0.57% - 52.77%, 1.29% - 50.82% for PA, 1.33% - 45.57%, 3.25% and 47.79 for PR. Regarding the explained variability (R^2), the EEMD-QRF-MOO model achieved better results than the other approaches. These improvements are ranged between 0.86% - 94.54%, 0.75% - 74.66% and 2.23% and 84.15% for PA, PR and SC states, respectively.

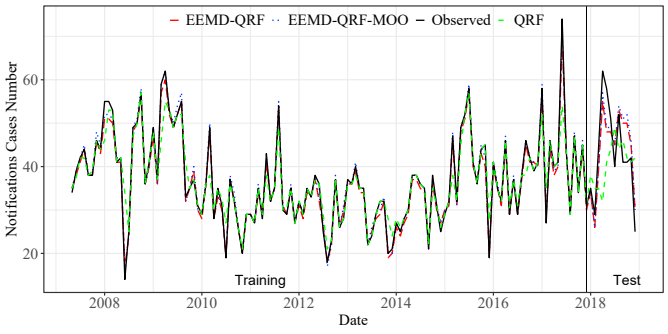
Additionally, results for SE standard deviation are presented in Table V. This measure is used to assess the models stability. These results show the stability of each model when predictions out-of-sample are evaluated. For all cases, the proposed approach shows better results than the other two approaches.

TABLE V
STANDARD DEVIATION FOR MODELS SE FOR TEST SET.

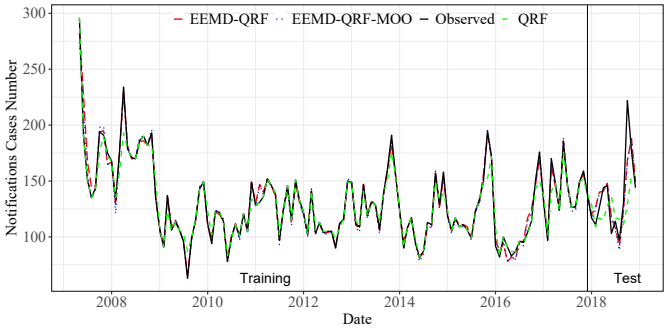
State	EEMD-QRF	EEMD-QRF-MOO	QRF
PA	1.0500	**0.9902**	6.0696
PR	2.1544	**2.1542**	7.2229
SC	1.6042	**1.4264**	4.3714

Figure 5a illustrates the observed versus predicted values for meningitis notified cases number obtained by EEMD-QRF, EEMD-QRF-MOO and QRF.

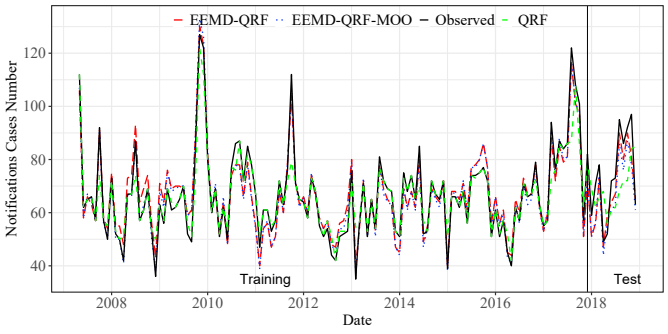
With regard of results obtained when EEMD associated or not with MOO techniques are employed, the forecasting are close of observed values for PA and PR states. Alongside,



(a) Observed and Predictions for PA state.



(b) Observed and Predictions for PR state.



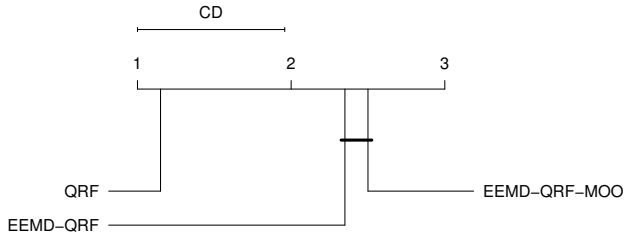
(c) Observed and Predictions for SC state

Fig. 5. Observed versus predicted values for the meningitis notified cases in each state.

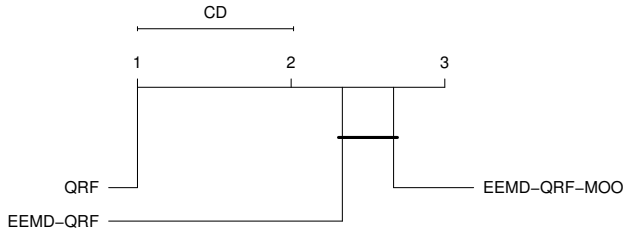
lower variations are observed. In contrast, greater error can be observed for forecasting values in SC state. However, even so these results are better than use only QRF approach. This suggest that employ signal decomposition allows to obtain forecast values similar with observed values than not using this approach. Additionally, adopting MOO for signal reconstruction enables to enhance the models performanceby dealing with the trade-off between the objectives defined in the MOP [40]. In other words, by the use of MOO approach, it is improved the models accuracy and stability by minimizing the forecast errors as well as variability of the errors. In this way, with the use of decomposition combined with MOO was possible deal of the nonstationarity of the data, and in the PR case, with the seasonality.

According to Friedman test, considering months that predictions are different, there is statistically difference between

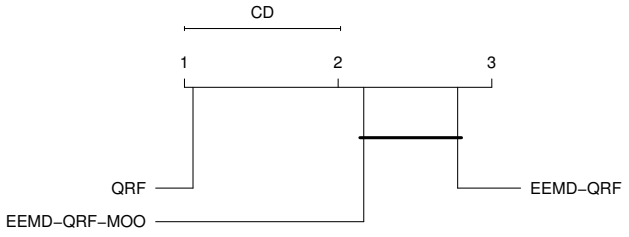
the SE for three adopted approaches. (χ^2_2 = 12.51 - 14.97, p-value < 0.05). Figure 6 illustrates the comparisons between three approaches. In this representation, those algorithms that are not joined by a line can be regarded as different. The CD to consider the errors statistically different are 1.1088, 1.1722 and 1.1722 for PA, PR and SC states with 2 degree of freedom.



(a) CD diagram for PA state.



(b) CD diagram for PR state.



(c) CD diagram for SC state

Fig. 6. Visualization of post-hoc test for SE of three approaches for PA, PR and SC states.

With regard to illustrated by Figure 6, indeed, the QRF approach presents statistically higher error than two approaches based on decomposition for all states (Difference = 1.11 - 1.72, p-value < 0.05). For PA and PR states, the EEMD-QRF-MOO shows lower error than EEMD-QRF, but there is no statistical difference (Difference = -0.33 - 0.61, p-value > 0.05). Even though no statistical difference is observed between the errors of the EEMD-QRF-MOO and EEMD-QRF approaches, this does not imply that the results obtained with the use of the former for PR and PA states and the latter for SC are the same. Given this, it appears that however small the differences exist and should be considered [41].

Considering the aforementioned results, the effectiveness of combined approach lies on two folds. First, the accuracy and stability of EEMD-QRF-MOO are reached due to the use of MOO, once this approach is designed to achieve the objectives of high accuracy and stability simultaneously [42]. Second, by EEMD use it is feasible to separate the internal characteristics of the original signal, which permit to improve the model's accuracy.

V. Conclusion

This paper proposed an ensemble based on decomposition EEMD, QRF and MOO (EEMD-QRF-MOO). The proposed framework was developed with objective to forecast one month ahead the meningitis notified cases number for PA, PR and SC States. This approach was compared with EEMD-QRF and QRF. The main difference of EEMD-QRF and EEMD-QRF-MOO is that in second approach the original signal is reconstructed using the MOO perspective. This allows to deal with bias-variance trade-off. The models were compared using the RMSE, sMAPE and R^2 criteria, as well as with statistical tests. According to the obtained results, it is concluded that:

1) Employing signal decomposition improve the final results for training and test sets regarding not apply decomposition;
2) The combination of signal decomposition, training each component with ensemble approach and reconstruct the signal using optimization enable to enhance the models performance for adopted time series;
3) Results from two signal reconstruction approaches are statistically equal, while the EEMD-QRF-MOO outperform EEMD-QRF in absolute values on 2 of 3 cases.

For future works is intend to adopt other decomposition techniques, training each component by other ensemble and single approaches, use other forecast horizons, MOO algorithms as well as different MCDM techniques.

Acknowledgments

The authors would like to thank National Council of Scientific and Technologic Development of Brazil – CNPq (Grants number: 303908/2015-7-PQ, 303906/2015-4-PQ, 404659/2016-0-Univ, 405101/2016-3-Univ), PRONEX '*Fundação Araucária*' 042/2018, and *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* - Brazil – CAPES (Finance Code: 321414/2019-0-PROSUC and 353671/2019-0-PROEX) for its financial support of this work. Furthermore, the authors would like to thank the Program Chairs and Reviewers for their constructive comments and recommendations, which have significantly improved the presentation of this paper.

References

[1] W. F. Wright, C. N. Pinto, K. Palisoc, and S. Baghli, "Viral (aseptic) meningitis: A review," *Journal of the Neurological Sciences*, vol. 398, pp. 176 – 183, 2019.

[2] I. Savin, K. Ershova, N. Kurdyumova, O. Ershova, O. Khomenko, G. Danilov, M. Shifrin, and V. Zelman, "Healthcare-associated ventriculitis and meningitis in a neuro-ICU: Incidence and risk factors selected by machine learning approach," *Journal of Critical Care*, vol. 45, pp. 95 – 104, 2018.

[3] D. Jiang, M. Hao, F. Ding, J. Fu, and M. Li, "Mapping the transmission risk of zika virus using machine learning models," *Acta Tropica*, vol. 185, pp. 391 – 399, 2018.

[4] J. M. Scavuzzo, F. Trucco, M. Espinosa, C. B. Tauro, M. Abril, C. M. Scavuzzo, and A. C. Frery, "Modeling dengue vector population using remotely sensed data and machine learning," *Acta Tropica*, vol. 185, pp. 167 – 175, 2018.

[5] M. Soliman, V. Lyubchich, and Y. R. Gel, "Complementing the power of deep learning with statistical model fusion: Probabilistic forecasting of influenza in Dallas County, Texas, USA," *Epidemics*, p. 100345, 2019.

[6] D. Yang and Z. Dong, "Operational photovoltaics power forecasting using seasonal time series ensemble," *Solar Energy*, vol. 166, pp. 529 – 541, 2018.

[7] E. L. Bullock, C. E. Woodcock, and C. E. Holden, "Improved change monitoring using an ensemble of time series algorithms," *Remote Sensing of Environment*, 2019.

[8] T. G. Dietterich, "Ensemble methods in machine learning," in *International Workshop on Multiple Classifier Systems*. Berlim, Germany: Springer, 2000, pp. 1–15.

[9] H. Liu, Y. Xu, and C. Chen, "Improved pollution forecasting hybrid algorithms based on the ensemble method," *Applied Mathematical Modelling*, vol. 73, pp. 473 – 486, 2019.

[10] G. T. Ribeiro, V. C. Mariani, and L. dos Santos Coelho, "Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting," *Engineering Applications of Artificial Intelligence*, vol. 82, pp. 272 – 281, 2019.

[11] P. Guo, Q. Zhang, Y. Chen, J. Xiao, J. He, Y. Zhang, L. Wang, T. Liu, and W. Ma, "An ensemble forecast model of dengue in Guangzhou, China using climate and social media surveillance data," *Science of The Total Environment*, vol. 647, pp. 752 – 762, 2019.

[12] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.

[13] J. Huan, W. Cao, and Y. Qin, "Prediction of dissolved oxygen in aquaculture based on EEMD and LSSVM optimized by the Bayesian evidence framework," *Computers and Electronics in Agriculture*, vol. 150, pp. 257–265, 2018.

[14] Y.-X. Wu, Q.-B. Wu, and J.-Q. Zhu, "Improved EEMD-based crude oil price forecasting using LSTM networks," *Physica A: Statistical Mechanics and its Applications*, vol. 516, pp. 114–124, 2019.

[15] W.-J. Niu, Z.-K. Feng, M. Zeng, B.-F. Feng, Y.-W. Min, C.-T. Cheng, and J.-Z. Zhou, "Forecasting reservoir monthly runoff via ensemble empirical mode decomposition and extreme learning machine optimized by an improved gravitational search algorithm," *Applied Soft Computing*, vol. 82, p. 105589, 2019.

[16] K. Vaysse and P. Lagacherie, "Using quantile regression forest to estimate uncertainty of digital soil mapping products," *Geoderma*, vol. 291, pp. 55 – 64, 2017.

[17] W. Zhang, H. Quan, and D. Srinivasan, "Parallel and reliable probabilistic load forecasting via quantile regression forest and quantile determination," *Energy*, vol. 160, pp. 810 – 819, 2018.

[18] N. Khan, S. Shahid, L. Juneng, K. Ahmed, T. Ismail, and N. Nawaz, "Prediction of heat waves in Pakistan using quantile regression forests," *Atmospheric Research*, vol. 221, pp. 1–11, 2019.

[19] Z. Qu, K. Zhang, W. Mao, J. Wang, C. Liu, and W. Zhang, "Research and application of ensemble forecasting based on a novel multi-objective optimization algorithm for wind-speed forecasting," *Energy Conversion and Management*, vol. 154, pp. 440–454, 2017.

[20] J. Zhu, J. Liu, P. Wu, H. Chen, and L. Zhou, "A novel decomposition ensemble approach to crude oil price forecasting with evolution clustering and combined model," *International Journal of Machine Learning and Cybernetics*, pp. 1–14, 2019.

[21] P. Du, J. Wang, Z. Guo, and W. Yang, "Research and application of a novel hybrid forecasting system based on multi-objective optimization for wind speed forecasting," *Energy Conversion and Management*, vol. 150, pp. 90–107, 2017.

[22] X. Zhang and J. Wang, "A novel decomposition ensemble model for forecasting shortterm load-time series with multiple seasonal patterns," *Applied Soft Computing*, vol. 65, pp. 478–494, 2018.

[23] Brasil, *Ministério da Saúde, Departamento de Informática do Sistema Único de Saúde (DATASUS)*, 2019, access in 22 jul. 2019, (in Portuguese). [Online]. Available: http://www2.datasus.gov.br/DATASUS/index.php?area=0203&id=29892234&VObj=http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sinannet/cnv/menin

[24] P. A. Morettin and C. Toloi, *Análise de séries temporais*, 2nd ed. So Paulo, Brasil: Blucher, 2006, (in Portuguese).

[25] S. Wang, N. Zhang, L. Wu, and Y. Wang, "Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and ga-bp neural network method," *Renewable Energy*, vol. 94, pp. 629–636, 2016.

[26] Z. Qian, Y. Pei, H. Zareipour, and N. Chen, "A review and discussion of decomposition-based hybrid models for wind energy forecasting applications," *Applied Energy*, vol. 235, pp. 939 – 953, 2019.

[27] N. Meinshausen, "Quantile regression forests," *Journal of Machine Learning Research*, vol. 7, pp. 983–999, 2006.

[28] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct 2001.

[29] R. Marler and J. Arora, "Survey of multi-objective optimization methods for engineering," *Structural and Multidisciplinary Optimization*, vol. 26, no. 6, pp. 369–395, 2004.

[30] G. Reynoso-Meza, X. Blasco, J. Sanchis, and M. Martínez, "Controller tuning using evolutionary multi-objective optimisation: Current trends and applications," *Control Engineering Practice*, vol. 28, pp. 58–73, 2014.

[31] J. Neville and D. Jensen, "Bias/variance analysis for relational domains," in *Inductive Logic Programming*, H. Blockeel, J. Ramon, J. Shavlik, and P. Tadepalli, Eds. Berlin, Germany: Springer, 2008, pp. 27–28.

[32] Y. Xiang, L. Gou, L. He, S. Xia, and W. Wang, "A SVR-ANN combined model based on ensemble EMD for rainfall prediction," *Applied Soft Computing*, vol. 73, pp. 874 – 883, 2018.

[33] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, April 2002.

[34] C.-L. Hwang and K. Yoon, *Methods for multiple attribute decision making*. Berlin, Germany: Springer, 1981, pp. 58–191.

[35] J. Derrac, S. García, D. Molina, and F. Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms," *Swarm and Evolutionary Computation*, vol. 1, no. 1, pp. 3–18, 2011.

[36] P. Nemenyi, "Distribution-free multiple comparisons," Ph.D. dissertation, Statistics Schoool, Princeton University, Princeton, USA, 1963.

[37] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *The Journal of Machine Learning Research*, vol. 7, pp. 1–30, Dec. 2006.

[38] S. García, A. Fernández, J. Luengo, and F. Herrera, "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power," *Information Sciences*, vol. 180, no. 10, pp. 2044 – 2064, 2010.

[39] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018. [Online]. Available: http://www.R-project.org/

[40] M. H. D. M. Ribeiro, V. H. A. Ribeiro, G. Reynoso-Meza, and 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, July 2019, pp. 1–8.

[41] D. G. Altman and J. M. Bland, "Statistics notes: Absence of evidence is not evidence of absence," *Bmj*, vol. 311, no. 7003, p. 485, 1995.

[42] J. Wang, W. Yang, P. Du, and T. Niu, "A novel hybrid forecasting system of wind speed based on a newly developed multi-objective sine cosine algorithm," *Energy Conversion and Management*, vol. 163, pp. 134 – 150, 2018.