COBEM 2021 Florianópolis - Brasil
ABCM Associação Brasileira de Engenharia e Ciências Mecânicas
26th ABCM International Congress of Mechanical Engineering
November 22-26, 2021. Florianópolis, SC, Brazil

COB-2021-0166
AN IMPROVED ENSEMBLE LEARNING MODEL FOR MULTI-STEP AHEAD WIND POWER GENERATION FORECASTING

Matheus Henrique Dal Molin Ribeiro[1,2]
mribeiro@utfpr.edu.br
Ramon Gomes da Silva[1]
gomes.ramon@pucpr.edu.br
Sinvaldo Rodrigues Moreno[3]
sinvaldo.moreno@gmail.com
Viviana Cocco Mariani[3,4]
viviana.mariani@pucpr.br
Leandro dos Santos Coelho[1,4]
leandro.coelho@pucpr.br

[1] Industrial and Systems Engineering Graduate Program (PPGEPS), Pontifical Catholic University of Parana (PUCPR). 1155, Rua Imaculada Conceição, Curitiba, PR, Brazil. 80215-901
[2] Department of Mathematics, Federal University of Technology - Parana (UTFPR). Via do Conhecimento, KM 01 - Fraron, Pato Branco, PR, Brazil. 85503-390
[3] Mechanical Engineering Graduate Program (PPGEM), Pontifical Catholic University of Parana (PUCPR). 1155, Rua Imaculada Conceição, Curitiba, PR, Brazil. 80215-901
[4] Department of Electrical Engineering, Federal University of Parana (UFPR). 100, Avenida Coronel Francisco Heraclito dos Santos, Curitiba, PR, Brazil. 81530-000

Abstract. *The development and expansion of clean energy, such as wind energy, are important in the preservation of the environment and development of local economies and an alternative to hydroelectric and thermal energies. In this respect, the development of efficient forecasting models to support the decision-making process is necessary. However, the effect of climatic and demographic factors makes it challenging. This study evaluates bootstrap aggregation efficiency (bagging) combined with a stacking ensemble learning model for short and medium-term (one up to twelve hours ahead) forecasting wind turbine wind power generation for a wind farm located in Parazinho, Brazil. The forecasting accuracy is evaluated through the root mean squared error, mean absolute error, and Theil's U index of inequality (type 2). The results suggest that for one-hour-ahead forecasting wind power generation, the stacking ensemble learning achieves forecasting errors lower than the combination of stacking with bagging ensemble approach according to all performance criteria and have competitive results concerning the remaining forecasting. In 85.42% of the comparisons, the stacking combined with the bagging ensemble has better accuracy than the stacking ensemble learning model regarding the adopted criteria.*

Keywords: *wind power, time series forecasting, bagging, stacking, machine learning*

1. INTRODUCTION

Wind energy is a renewable, non-polluting, and clean energy resource, which the generation increase over the last years in Brazil. Then, forecasting wind power generation is key to developing correct strategic planning in energy power distribution. In this context, wind energy has been increasing its operation in the energy matrix in the last decades in many countries worldwide. According to the report of "INFOVENTO" Brazilian Wind Energy Association (ABEEólica) (2021), the size of the wind industry of Brazil in terms of installed capacity is 18.62 gigawatts produced by 713 wind farms in 12 states. Due to the high installed capacity of wind generation, 57 terawatt-hours of wind energy were generated in 2020, benefiting 86.4 million Brazilians. Economically, 35.8 billion dollars have already been invested in this area between 2011 and 2020. For every megawatt-hour of generation, 15 citizens are employed. Environmentally, 21.2 million tons of carbon dioxide (CO_2) were avoided in 2020. Wind power generation is dependent and influenced by climatic and demographic (da Silva *et al.*, 2021), as well as wind farms layout (Chowdhury *et al.*, 2012). Concerning it, the development of accurate forecasting models for wind power generation is challenging. Also, the erratic behavior and high wind speed changes with no typical patterns heavily make the forecasting process hard (Hossain *et al.*, 2021). To obtain an efficient forecasting method in terms of accuracy, the modeling of these features is essential.

Different studies have already been developed to obtain efficient forecasting frameworks for renewable energy contexts, especially for wind power generation. Fraccanabbia *et al.* (2020) combined the stacking ensemble learning model with different feature selection approaches to perform solar power forecasting. As a result, the authors argued that using a correlation matrix combined with a stacking framework could outperform other compared methods and lead to better forecasting results. da Silva *et al.* (2021) proposed an ensemble learning model by combining time series pre-processing, feature engineering, and stacking ensemble learning for wind power forecasting concise term. The proposed framework outperformed pre-processed ensemble learning models and single models with a performance improvement of forecasting errors that ranges 0.06% - 97.53%. González-Sopeña *et al.* (2021) evaluated the performance of decomposition-based hybrid models under different conditions in the context of wind power forecasting. The forecasting horizons of six and twenty-four hours ahead were adopted. The authors argued that depending on the kind of forecasting approach, such as point or interval (probabilistic), a set of metrics is better than others and the forecasting strategy (recursive or direct, for example). Shahid *et al.* (2021) proposed a novel genetic long short term memory (GLSTM) framework comprising of long short term memory and genetic algorithm is proposed to predict short-term wind power. Prediction from GLSTM has been compared with actual power, predictions of support vector regression, and reporting techniques in standard performance indices, such as mean absolute error (MAE), root mean squared error (RMSE), and mean squared error (MSE). In terms of percentage improvement, GLSTM, on average, improves wind power predictions from 6% to 30% instead of existing techniques. Putz *et al.* (2021) proposed a deep neural architecture for short-term wind power forecasting. In terms of percentage errors, the authors observed that as the forecasting horizon increases, forecasting errors also increase.

Regarding the aforementioned related works, it is evident the great attention give to the development of efficient forecasting models in the context of wind power generation. Also, due to the importance of renewable energy for the economy and environment in the national aspect, forecasting wind power generation as accurately as possible is necessary. Therefore, this study evaluates bootstrap aggregation efficiency (bagging) combined with a stacking ensemble learning model for short and medium-term (one up to twelve hours ahead) forecasting of wind power generation a wind turbine for a wind farm located in Parazinho city, State of Rio Grande do Norte, Brazil. In this approach, we aim to identify the suitable number of samples (10, 30, 50, or 100) obtained through bootstrap to compose the bagging model; and obtain the best strategy used in the bagging model, average or median. The forecasting accuracy is evaluated through the mean absolute error MAE, RMSE, normalized RMSE, and Theil's U index of inequality (UT).

The main contributions of this study can be summarized as follows: The first contribution is related to evaluating the use of a combined framework that integrates bagging and stacking ensemble learning methods for wind power generation forecasting. Second, the use of settings (number of bootstrap samples) and aggregation methods (average and median). Last, this study evaluates the proposed framework forecasting in a multi-step ahead forecasting strategy considering short (one up to six-hours-ahead) and medium-term (seven up to twelve-hours-ahead) horizons for time series in the renewable energy context.

The rest of this paper is structured as follows. Section 2 describes the dataset employed for analysis. Section 3 defines the adopted methods. Section 4 presents the results obtained and the discussions. Finally, Section 5 concludes with the final considerations and future works.

2. DATASET DESCRIPTION

In this study is adopted one time series composed by the wind power generation in kiloWatts-hour (kWh) recorded every 60 minutes, from one wind farm located in Parazinho city, State of Rio Grande do Norte, Brazil. The wind farm is composed of fifteen aero generators, each one with 2000 kWh of power, rotor diameter 90 m rotor height 80 m. The length of the time series reached 2,406 samples from August, 1st to November, 20th, 2020. with a sampling rate of one hour, as presented in Fig. 1.

Tab. 1 presents the statistical indicators of adopted time series.

Table 1: Statistical indicators of wind power generation time series.

Dataset	Statistical Indicator					
	Minimum	Median	Average	Maximum	Standard Deviation	# Samples
Whole	810.88	20093.76	19079.59	29627.22	6889.61	2406
Training	2108.69	22335.07	21100.39	29627.22	6000.78	1684
Test	810.88	14200.27	14366.25	29146.23	6518.65	722

3. METHODS

This section presents the main aspects of the methods proposed in this study. The bagging and stacking ensemble methods are presented followed by the description performance measures used in the evaluation of compared models.

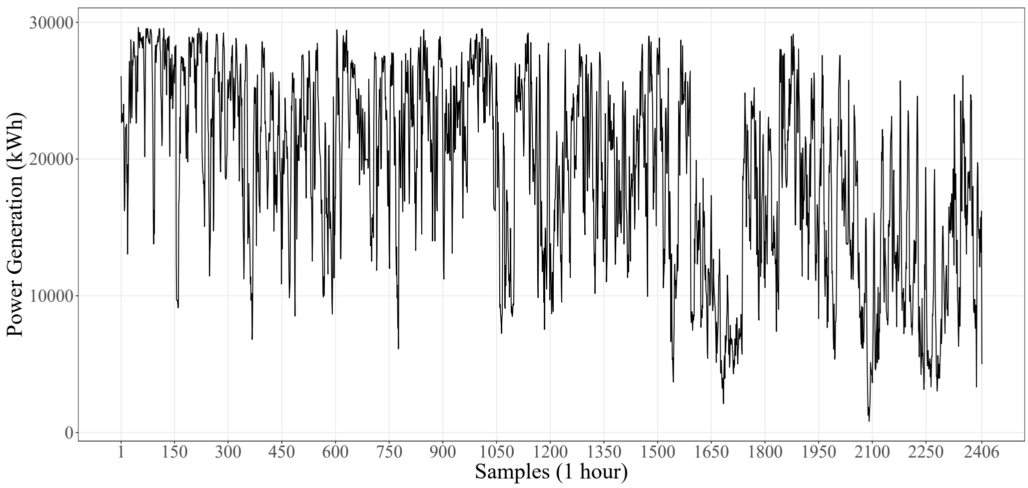


Figure 1: Wind power generation time series.

3.1 BAGGING ENSEMBLE LEARNING

Bagging, is a classical technique for the creation of an ensemble learning model proposed by Breiman (1996). It was initially proposed for use in classification problems. However, it can be used in problems that aim to perform data regression. It is characterized by the creation of multiple samples, with refitting employing the bootstrap technique, from the same set of data, so that it is possible to build multiple distinct trees for the same predictor and use them to generate an aggregate prediction. The final prediction from this process can be obtained by voting or average, for classification and regression problems respectively (Erdal and İlhami Karahanoğlu, 2016). This allows the generation of multiple samples for the same set. The advantage of using this technique for ensemble generation is that it allows for the reduction of error in the baseline predictors, which can be considered unstable before certain perturbations, and that it can provide its estimates of predictive performance, which correlate with estimates of cross-validation or estimates of test sets (Hamze-Ziabari and Bakhshpoori, 2018).

In this paper, the bagging strategy proposed by Bergmeir *et al.* (2016) is adopted. However, instead of using the exponential smoothing technique to deal with the different samples obtained through the bagging paradigm, we used the stacking ensemble learning method. The time series analysis is difficult once data are autocorrelated. In this case, instead of using a traditional bootstrap strategy for non-correlated data, the moving block bootstrap (MBB) is used. The MBB resamples the data inside overlapping blocks to imitate the autocorrelation in the data. The steps proposed by Bergmeir *et al.* (2016) are (i) the seasonal and Trend decomposition using Locally-weighted regression (STL) is applied in the data; (ii) the moving block bootstrapping is used for the remainder component, and next the original signal is reconstructed by summing up the bootstrapped remainder, trend, and seasonal components; and (iii) the stacking ensemble learning is applied in each sample to obtain the forecasting results. In this paper, the combination of different number of bootstrap samples, i.e., 10, 30, 50, and 100 samples are considered, and two aggregation strategies, i.e., average and median values.

3.2 STACKING ENSEMBLE LEARNING

The stacked generalized is an ensemble learning model based on the principle of divide-to-conquer (Ribeiro and Coelho, 2020), which improves models' accuracy by integrating models through layers that have been successfully applied to solve problems in several fields of knowledge. Usually, two layers are commonly adopted, however, it may not be limited to that. Indeed, the stacking ensemble learning is an efficient tool to deal with a different kind of time series features in some fields, such as financial (da Silva *et al.*, 2020), agriculture (Wu *et al.*, 2021), fault detection (Li *et al.*, 2021), image analysis (Anoop *et al.*, 2020).

In the first layer (layer-0), base-learners (weak-learner, or weak-models) are trained and its predictions are used in the next layer. In the sequence, for layer-1 a meta-learner (strong-learner, or strong-model) is trained, whose predictions of the previous layer are adopted as system inputs and predictions are obtained. In this paper, the base models are Gaussian processes and Support Vector Regression with linear kernel, k-Nearest Neighbor, and Random Forest. The meta-model is ridge regression.

3.3 PERFORMANCE MEASURES AND HYPOTHESIS TEST

To evaluate the effectiveness of adopted models, from obtained forecasts out-of-sample (test set), MAE, RMSE, normalized by average RMSE (nRMSE), and UT (type 2) performance criteria, are computed as follows,

$$\text{MAE} = \sum_{i=1}^{n} \frac{1}{n}|y_i - \hat{y}_i|, \tag{1}$$

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

$$\text{nRMSE} = 100 \times \frac{RMSE}{\bar{\mathbf{y}}}, \tag{3}$$

$$\text{UT} = \frac{\sqrt{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}}{\sqrt{\sum_{i=1}^{n} y_i^2}} \tag{4}$$

where n is the number of observations (samples), y_i and $\hat{y}_i$ are the *i*-th observed and predicted values, respectively, and $\bar{\mathbf{y}}$ is the average of observed values.

In this study, the Friedman test is used to verify if at least two of the models represents different results. The statistic of the test is stated according to,

$$FD = \frac{12n}{k(k+1)}\left[\sum_j R_j^2 - \frac{k(k+1)^2}{4}\right] \sim \chi^2_{k-1}, \tag{5}$$

where is distributed according to chi-squared distribution with $k-1$ degrees of freedom, with n observations and k groups, R_j^2 is the squared rank of j-th compared approach. Under the null hypothesis, there is no difference between results from different groups.

In the sequence, according to Carrasco *et al.* (2020), if Friedman's null hypothesis is rejected, it is necessary to apply a posthoc test to find which groups have different results. Hence, a multiple comparison test of Nemenyi can be applied. In this approach, a threshold is obtained by

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

where is distributed according to chi-squared distribution with $k-1$ degrees of freedom, with n observations and k groups, R_j^2 is the squared rank of j-th compared approach. Under the null hypothesis, there is no difference between results from different groups.

In the sequence, according to Carrasco *et al.* (2020), if Friedman's null hypothesis is rejected, it is necessary to apply a post-hoc test to find which groups have different results. Hence, a multiple comparison test of Nemenyi (Nemenyi, 1962) can be applied.

4. RESULTS AND DISCUSSIONS

This section describes the main results obtained by the proposed framework forecasting model for wind power forecasting. The performance metrics of the compared models are presented in Tab. 2 in some forecasting horizons (1 up to 12-hours-ahead). The best performance accuracy, regarding each one of the adopted forecasting metric for each model, are highlighted in bold.

In the context of wind power generation one-hour-ahead forecasting, the stacking ensemble learning, considered in the stand-alone framework, outperforms the compared methods in terms of RMSE, MAE, and UT. The improvement percentage is 13.42%, 13.82%, 12.94%, and 12.50% regarding the stacking combined with bagging, when 10, 30, 50,

Table 2: Performance measures of compared models for wind power generation forecasting.

Forecasting Horizon	Criteria	Stack	Stack + 10 Bootstraps		Stack + 30 Bootstraps		Stack + 50 Bootstraps		Stack + 100 Bootstraps	
		Original	Average	Median	Average	Median	Average	Median	Average	Median
1	RMSE	**2828.28**	3263.56	3269.56	3282.41	3281.12	3249.62	3247.45	3234.27	3230.37
	nRMSE	**19.69%**	22.72%	22.76%	22.85%	22.84%	22.62%	22.60%	22.51%	22.49%
	MAE	**2143.38**	2658.33	2654.99	2666.80	2667.84	2650.65	2645.63	2624.77	2616.27
	UT	**0.1757**	0.2088	0.2082	0.2104	0.2096	0.2084	0.2073	0.2073	0.2061
2	RMSE	3619.77	3627.37	3633.27	3619.22	3601.62	3615.68	3614.11	3589.65	**3587.29**
	nRMSE	25.20%	25.25%	25.29%	25.19%	25.07%	25.17%	25.16%	24.99%	**24.97%**
	MAE	**2794.92**	2948.79	2953.97	2950.43	2935.89	2953.71	2949.19	2927.37	2919.55
	UT	**0.2228**	0.2309	0.2305	0.2309	0.2293	0.2308	0.2296	0.2290	0.2279
3	RMSE	3965.14	3843.91	3831.81	3847.81	3832.02	3851.32	3850.22	3819.64	**3816.36**
	nRMSE	27.60%	26.76%	26.67%	26.78%	26.67%	26.81%	26.80%	26.59%	**26.56%**
	MAE	**3106.56**	3150.51	3149.94	3133.94	3112.65	3156.10	3149.30	3121.94	3116.57
	UT	0.2442	0.2446	0.2431	0.2457	0.2443	0.2456	0.2448	0.2436	**0.2426**
4	RMSE	4299.16	4038.70	4020.25	4036.19	4021.39	4031.32	4024.26	4017.22	**4015.05**
	nRMSE	29.93%	28.11%	27.98%	28.09%	27.99%	28.06%	28.01%	27.96%	**27.95%**
	MAE	3439.89	3307.92	3306.61	3281.34	**3268.64**	3303.48	3291.35	3287.81	3285.06
	UT	0.2655	0.2571	0.2558	0.2574	0.2561	0.2576	0.2559	0.2563	**0.2554**
5	RMSE	4745.92	4311.46	4298.85	4253.71	**4237.61**	4262.42	4267.57	4275.31	4279.48
	nRMSE	33.04%	30.01%	29.92%	29.61%	**29.50%**	29.67%	29.71%	29.76%	29.79%
	MAE	3753.23	3563.41	3553.66	3485.50	**3472.47**	3524.86	3523.52	3523.55	3519.53
	UT	0.2857	0.2705	0.2694	0.2679	**0.2666**	0.2676	0.2672	0.2679	0.2673
6	RMSE	4980.08	4431.71	4422.91	4407.80	4396.45	4405.82	4414.56	**4380.39**	4381.94
	nRMSE	34.67%	30.85%	30.79%	30.68%	30.60%	30.67%	30.73%	**30.49%**	30.50%
	MAE	4068.44	3629.18	3630.24	3600.70	3583.04	3613.53	3615.79	3588.88	**3580.12**
	UT	0.3048	0.2845	0.2830	0.2844	0.2834	0.2834	0.2831	0.2813	**0.2807**
7	RMSE	5270.26	4617.50	4628.60	**4533.12**	4542.79	4580.91	4594.23	4581.73	4559.18
	nRMSE	36.69%	32.14%	32.22%	**31.55%**	31.62%	31.89%	31.98%	31.89%	31.74%
	MAE	4330.83	3815.19	3820.15	**3715.38**	3724.50	3770.34	3772.40	3756.30	3738.98
	UT	0.3090	0.2888	0.2884	**0.2827**	0.2831	0.2856	0.2858	0.2855	0.2836
8	RMSE	5642.22	4944.71	4947.58	4886.27	4876.31	4885.22	4893.83	**4871.47**	4881.43
	nRMSE	39.27%	34.42%	34.44%	34.01%	33.94%	34.00%	34.06%	**33.91%**	33.98%
	MAE	4623.75	4078.54	4088.49	**3976.37**	3977.77	4012.46	4013.49	4001.76	4010.42
	UT	0.3342	0.3080	0.3072	0.3073	0.3062	0.3060	0.3054	0.3049	**0.3046**
9	RMSE	5630.98	5013.70	5027.65	**4976.59**	4989.36	5008.38	5020.91	5021.83	5018.92
	nRMSE	39.20%	34.90%	35.00%	**34.64%**	34.73%	34.86%	34.95%	34.96%	34.94%
	MAE	4572.41	4131.90	4157.31	**4102.29**	4107.94	4132.87	4129.94	4139.83	4131.96
	UT	0.3334	0.3106	0.3105	0.3103	**0.3101**	0.3110	0.3110	0.3122	0.3111
10	RMSE	5672.43	5088.86	5110.29	**5022.55**	5029.13	5042.02	5058.06	5033.94	5038.41
	nRMSE	39.48%	35.42%	35.57%	**34.96%**	35.01%	35.10%	35.21%	35.04%	35.07%
	MAE	4565.26	4213.07	4237.94	**4134.96**	4142.64	4171.57	4183.56	4159.95	4162.16
	UT	0.3336	0.3136	0.3145	0.3116	0.3116	0.3105	0.3105	0.3106	**0.3100**
11	RMSE	5851.22	5206.03	5193.34	**5132.27**	5137.94	5146.48	5162.33	5150.96	5167.22
	nRMSE	40.73%	36.24%	36.15%	**35.72%**	35.76%	35.82%	35.93%	35.85%	35.97%
	MAE	4703.26	4275.71	4266.79	**4194.70**	4206.75	4227.47	4238.53	4217.68	4229.88
	UT	0.3352	0.3201	0.3175	0.3177	0.3172	0.3158	**0.3158**	0.3166	0.3166
12	RMSE	6004.80	5419.08	5385.52	5469.57	5473.68	**5380.41**	5404.83	5410.85	5424.00
	nRMSE	41.80%	37.72%	37.49%	38.07%	38.10%	**37.45%**	37.62%	37.66%	37.76%
	MAE	4968.50	4470.93	4451.70	4486.47	4484.83	**4413.89**	4436.98	4446.02	4451.88
	UT	0.3569	0.3424	**0.3397**	0.3476	0.3480	0.3410	0.3413	0.3424	0.3423

and 100 bootstrap samples are considered, respectively. When the MAE and UT criteria are adopted, similar results are achieved, independently of the aggregation strategy adopted in bagging ensemble learning. Considering the second forecasting horizon, the stacking ensemble achieves better results than compared models in terms of MAE and UT. Also, three hours ahead, it achieves best results only when the MAE criterion is adopted, and worse results than stacking combined with bagging (100 samples) regarding RMSE and UT. Indeed, for wind energy, the UT meaning an equal distribution, while higher values meaning a higher level of inequality regarding the observed and forecasting wind power.

Concerning the remaining forecasting horizons, four, five, and seven to eleven hours ahead, the stacking combined with bagging (30 samples) outperforms compared models, in most of the indicators, in 41.67% (27.78% and 13.89% using average and median aggregation strategies, respectively) of the comparisons. In respect to fours-hours-ahead, in terms of MAE, the improvement percentage of bagging stacking ensemble learning regarding the remaining models range between 0.54% and 4.98%. The best and worse improvements are achieved in comparison with stacking and bagging stacking with

100 samples, respectively. For six hours ahead, the bagging stacking with 100 samples can reach better results than other considered forecasting structures. In terms of RMSE, the best aggregation strategy is the average, while for MAE and UT, the median. In fact, the improvement ranges between 0.49% - 12.04%, 0.33% - 12%, and 0.89% - 7.90%, for median, in terms of RMSE, MAE, and UT, respectively. Tab. 3 details the frequency distribution of best results for the compared models, according to the results presented in Tab. 2. It is possible to observe that the combination of bagging and stacking can achieve the best results in 83.33% of the comparisons. Also, the use of average or median aggregation reaches similar results in terms of frequency of best results considering the MAE, RMSE, and UT criteria, respectively. Concerning the bootstrap samples, better forecasting results regarding the performance measures are achieved when 30 bootstrap samples are considered in the ensemble structure, followed by 100, 50, and 10.

Table 3: Frequency distribution of best results according to the compared methods.

Aggregation	Stacking	Stacking + 10 Bootstraps	Stacking +30 Bootstraps	Stacking +50 Bootstraps	Stacking + 100 Bootstraps	Total
-	14.58%	-	-	-	-	14.58%
Average	-	0.00%	29.17%	6.25%	8.33%	43.75%
Median	-	2.08%	12.50%	2.08%	25%	41.67%
Total	14.58%	2.08%	41.67%	8.33%	33.33%	100.00%

In Fig. 2 are showed the observed and wind power forecasting for one up twelve hours ahead. For the first five forecasting horizons, the combination of stacking ensemble with bagging learns data behavior, in most of the cases, which allows predictions compatible with the observed values. There is a challenge for the remaining forecasting horizons. Moreover, in the ups and downs, the forecasting model has difficulty in capturing the data variability. However, usually, the obtained forecasting results can be useful to develop the decision-making process in the management of wind farms.

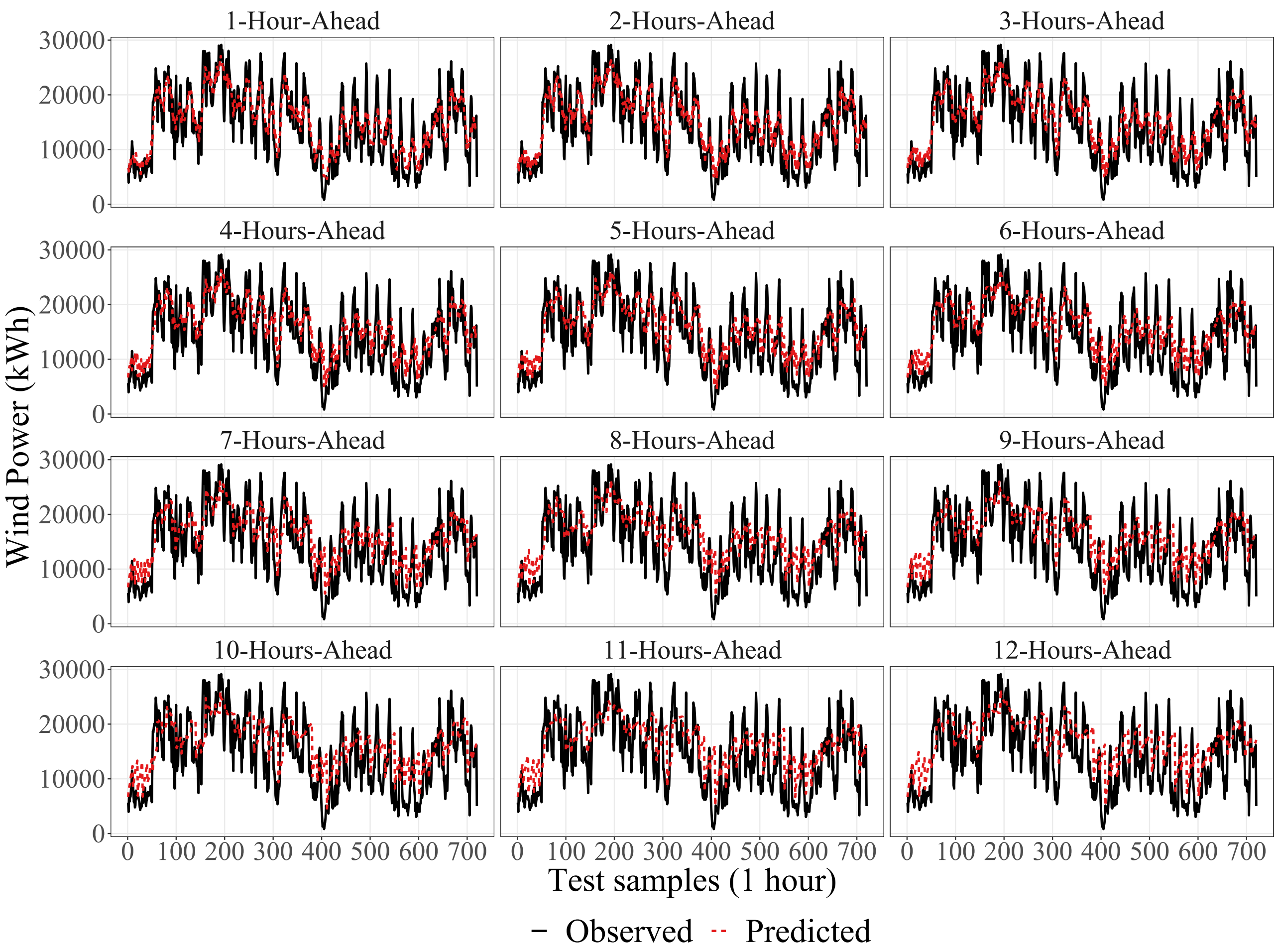
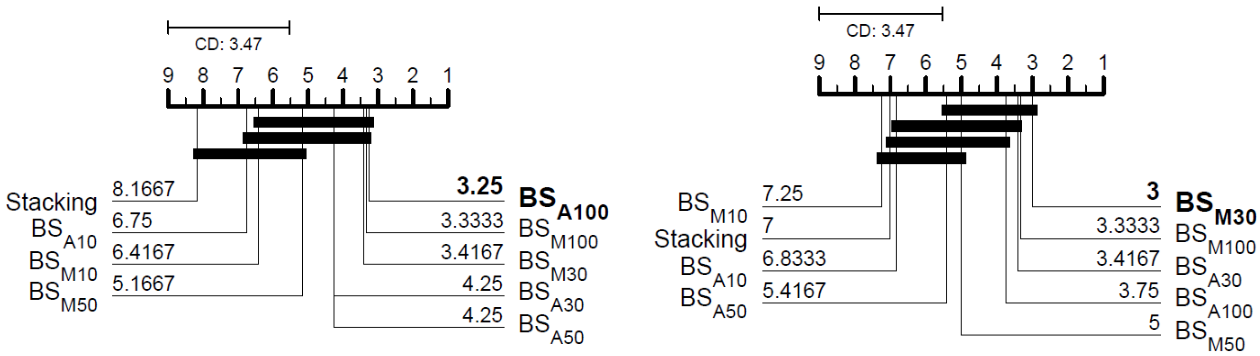


Figure 2: Observed versus predicted time series all forecasting horizons.

Considering Friedmann's test of the RMSE and MAE overall forecasting horizons for the compared methods, there are statistically different results to at least two models at 5% level ($\chi^2_8 = 39.35 \times 10^{-6}$ - 9.26×10^{-6}, p-value < 0.05).



(a) Diagram of CD for RMSE. (b) Diagram of CD for MAE.

Figure 3: Diagram of CD for post-hoc Nemenyi test for the evaluated models over all forecasting horizons.

Fig. 3a and 3b show the CD plot based on Nemenyi test (Nemenyi, 1962), and the compared models are stacking, bagging (average aggregation with 10 bootstraps) combined with stacking (BS_{A10}), bagging (average aggregation with 10, 30, 50, and 100 bootstraps) combined with stacking ($BS_{A10}, BS_{A30}, BS_{A50}, BS_{A100}$), and (median aggregation with 10, 30, 50, and 100 bootstraps) combined with stacking ($BS_{M10}, BS_{M30}, BS_{M50}, BS_{M100}$). Those forecasting models that are not joined by a line can be regarded as different. The CD to consider the errors statistically different is 3.47 in both cases. In both cases, the combination of bagging and stacking achieved the best ranking.

5. CONCLUSION AND FUTURE RESEARCH

In this study, machine learning approaches named bagging and stacking were combined to develop the task of forecasting one up to twelve hours ahead of the wind power generation of a wind farm located at Parazinho city, Brazil. The bagging ensemble is used to sub-sample (10, 30, 50, and 100 samples) the residual component obtained from STL decomposition, using MBB. Each sub-sample was trained and fitted with stacking ensemble learning models. The forecasting of the sub-samples was grouped through average and median to generate final forecasting models. The MAE, RMSE, and UT criteria were adopted to evaluate the performance of the compared approaches.

In a broader perspective, in 85.42% of the comparisons, the stacking combined with bagging ensemble has better accuracy than the stacking ensemble learning model regarding the performance measures. The results suggest that for one-hour-ahead forecasting wind power generation, the stacking ensemble learning achieves forecasting errors lower than the combination of stacking with bagging ensemble approach according to all performance criteria. Also, these two approaches have competitive results concerning the forecasting horizons of two and three-hours-ahead. When dealing with the number of bootstrap samples, better forecasting results regarding the performance measures are achieved when 30 bootstrap samples are considered in the ensemble structure, followed by 100, 50, and 10. Regarding the aggregation strategy, the average and median approaches provide similar forecasting accuracy regarding the adopted criteria.

For future works, it is intended to adopt (i) exogenous variables to try to improve the forecasting results and giving additional information for the forecasting system, (ii) to integrate the bagging with different time-series pre-processing, and (iii) optimization approach to select the base models of stacking ensemble learning model.

6. ACKNOWLEDGEMENTS

The authors would like to thank the National Council of Scientific and Technological Development of Brazil – CNPq (Grants number: 307958/2019-1-PQ, 307966/2019-4-PQ, 404659/2016-0-Univ, 160817/2019-6-PDJ), PRONEX '*Fundação Araucária*' 042/2018, and *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil* (CAPES) - Finance Code 001 for financial support of this work.

7. REFERENCES

Anoop, B., Pavan, R., Girish, G., Kothari, A.R. and Rajan, J., 2020. "Stack generalized deep ensemble learning for retinal layer segmentation in optical coherence tomography images". *Biocybernetics and Biomedical Engineering*, Vol. 40, No. 4, pp. 1343–1358. ISSN 0208-5216. doi:10.1016/j.bbe.2020.07.010.

Bergmeir, C., Hyndman, R. and Benítez, J., 2016. "Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation". *International Journal of Forecasting*, Vol. 32, No. 2, pp. 303–312. ISSN 0169-2070. doi:10.1016/j.ijforecast.2015.07.002.

Brazilian Wind Energy Association (ABEEólica), 2021. "Energia eólica: os bons ventos do Brasil - INFOVENTO 20".

URL `http://abeeolica.org.br/wp-content/uploads/2021/06/2021_05_InfoVento20.pdf`. (accessed 10 June, 2021, in Portuguese).

Breiman, L., 1996. "Bagging predictors". *Machine Learning*, Vol. 24, No. 2, pp. 123–140. ISSN 1573-0565. doi: 10.1023/A:1018054314350.

Carrasco, J., García, S., Rueda, M., Das, S. and Herrera, F., 2020. "Recent trends in the use of statistical tests for comparing swarm and evolutionary computing algorithms: Practical guidelines and a critical review". *Swarm and Evolutionary Computation*, Vol. 54, No. 100665. ISSN 2210-6502. doi:10.1016/j.swevo.2020.100665.

Chowdhury, S., Zhang, J., Messac, A. and Castillo, L., 2012. "Unrestricted wind farm layout optimization (UWFLO): Investigating key factors influencing the maximum power generation". *Renewable Energy*, Vol. 38, No. 1, pp. 16–30. ISSN 0960-1481. doi:10.1016/j.renene.2011.06.033.

da Silva, R.G., Ribeiro, M.H.D.M., Fraccanabbia, N., Mariani, V.C. and Coelho, L.S., 2020. "Multi-step ahead bitcoin price forecasting based on VMD and ensemble learning methods". In *International Joint Conference on Neural Networks (IJCNN)*. IEEE, Glasgow, United Kingdom, pp. 1–8. doi:10.1109/ijcnn48605.2020.9207152.

da Silva, R.G., Ribeiro, M.H.D.M., Moreno, S.R., Mariani, V.C. and Coelho, L.S., 2021. "A novel decomposition-ensemble learning framework for multi-step ahead wind energy forecasting". *Energy*, Vol. 216, No. 119174. doi: 10.1016/j.energy.2020.119174.

Erdal, H. and İlhami Karahanoğlu, 2016. "Bagging ensemble models for bank profitability: An emprical research on turkish development and investment banks". *Applied Soft Computing*, Vol. 49, pp. 861 – 867. ISSN 1568-4946. doi:10.1016/j.asoc.2016.09.010.

Fraccanabbia, N., da Silva, R.G., Moreno, S.R., Ribeiro, M.H.D.M., Mariani, V.C. and Coelho, L.S., 2020. "Solar power forecasting based on ensemble learning methods". In *IEEE World Congress on Computational Intelligence (IEEE WCCI), International Joint Conference on Neural Networks (IJCNN)*. Glasgow, Scotland, pp. 1–8. doi: 10.1109/IJCNN48605.2020.9206777.

González-Sopeña, J., Pakrashi, V. and Ghosh, B., 2021. "An overview of performance evaluation metrics for short-term statistical wind power forecasting". *Renewable and Sustainable Energy Reviews*, Vol. 138, No. 110515. ISSN 1364-0321. doi:10.1016/j.rser.2020.110515.

Hamze-Ziabari, S. and Bakhshpoori, T., 2018. "Improving the prediction of ground motion parameters based on an efficient bagging ensemble model of M5 and CART algorithms". *Applied Soft Computing*, Vol. 68, pp. 147–161. ISSN 1568-4946. doi:10.1016/j.asoc.2018.03.052.

Hossain, M.A., Chakrabortty, R.K., Elsawah, S. and Ryan, M.J., 2021. "Very short-term forecasting of wind power generation using hybrid deep learning model". *Journal of Cleaner Production*, Vol. 296, No. 126564. ISSN 0959-6526. doi:10.1016/j.jclepro.2021.126564.

Li, G., Zheng, Y., Liu, J., Zhou, Z., Xu, C., Fang, X. and Yao, Q., 2021. "An improved stacking ensemble learning-based sensor fault detection method for building energy systems using fault-discrimination information". *Journal of Building Engineering*, , No. 102812. ISSN 2352-7102. doi:10.1016/j.jobe.2021.102812.

Nemenyi, P., 1962. "Distribution-free multiple comparisons". In *Biometrics*. International Biometric Society, Vol. 18, p. 263.

Putz, D., Gumhalter, M. and Auer, H., 2021. "A novel approach to multi-horizon wind power forecasting based on deep neural architecture". *Renewable Energy*, Vol. 178, pp. 494–505. ISSN 0960-1481. doi:10.1016/j.renene.2021.06.099.

Ribeiro, M.H.D.M. and Coelho, L.S., 2020. "Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series". *Applied Soft Computing*, Vol. 86, No. 105837. doi: 10.1016/j.asoc.2019.105837.

Shahid, F., Zameer, A. and Muneeb, M., 2021. "A novel genetic lstm model for wind power forecast". *Energy*, Vol. 223, No. 120069. ISSN 0360-5442. doi:10.1016/j.energy.2021.120069.

Wu, T., Zhang, W., Jiao, X., Guo, W. and Alhaj Hamoud, Y., 2021. "Evaluation of stacking and blending ensemble learning methods for estimating daily reference evapotranspiration". *Computers and Electronics in Agriculture*, Vol. 184, No. 106039. ISSN 0168-1699. doi:10.1016/j.compag.2021.106039.

8. RESPONSIBILITY NOTICE

The authors are solely responsible for the printed material included in this paper.