Short-Term Wind Power Forecasting by a Long Short Term Memory Ensemble Approach

Geovanny Marulanda
Institute for Research in Technology (IIT)
ICAI School of Engineering, Comillas Pontifical University
Madrid, Spain
geovanny.marulanda@iit.comillas.edu

Jenny Cifuentes
ICADE, Faculty of Economics and Business Administration; Department of Quantitative Methods
Universidad Pontificia Comillas,
28015, Madrid, Spain
jacifuentes@comillas.edu

Antonio Bello
Institute for Research in Technology (IIT)
ICAI School of Engineering, Comillas Pontifical University
Madrid, Spain
antonio.bello@iit.comillas.edu

Javier Reneses
Institute for Research in Technology (IIT)
ICAI School of Engineering, Comillas Pontifical University
Madrid, Spain
javier.reneses@iit.comillas.edu

***Abstract*— Wind power plants have attracted increasing attention during the last decades due to their environmental and economic benefits. However, the wind resource is inherently unpredictable, bringing important challenges to the stable and safe operation of the power grid. In this context, various computational and statistical approaches have been reported in the literature to perform short-term forecasting of wind power generation, and more efficient strategies are still demanded. In this paper, a hybrid framework that includes a statistical pre-processing stage with an enhanced deep learning (DL)-based strategy is proposed to address the limitations of reported forecasting methodologies to predict multi-seasonal wind power time series. The integrated approach applies a suitable transformation to obtain a normal distribution of data and removes multiple seasonalities in wind power time series. Subsequently, it supplements a set of stacked Long Short Term Memory (LSTM) Recurrent Neural Network (RNN) models for each month of the year. The proposed approach is validated using real hourly wind power data from the Spanish electricity market for the period 2008-2019. A comparative analysis with a well-established DL-based model shows the superior performance of the proposed forecasting method. The experimental evaluation is conducted for 1-3 hours ahead of wind power predictions.**
Keywords—Long Short Term Memory; Deep Learning; Wind Power Forecasting; Recurrent Neural Networks; Time Series Decomposition

I. INTRODUCTION

Wind power generation has drawn an increasing interest since it has been recognized as a promissory renewable energy alternative to supplement traditional fossil energy. Although wind power has become one of the fastest-increasing energy sources, its variability and uncertainty bring important challenges to the stable and safe operation of the power grid [1], [2].

In this regard, short-term forecasting has been considered a critical approach to solving the problem. As such, the accurate estimation of wind power can improve its utilization, increase the reliability of the power system, reduce operation costs and enhance the development of efficient load management schemes. In addition, from the firms' point of view, wind power prediction is useful to mitigate the risk exposure and design energy portfolios for the short-term market.

Different strategies have been developed in this research field during the last few years. In particular, wind power forecasting strategies can be divided into three categories based on their modeling approach: physical methods, statistical and soft computing algorithms [3]. Physical methods mainly establish a prediction model based on a set of physical laws which characterize the corresponding meteorological process [4]. These approaches tend to be very complex and strongly dependent on the analyzed location. Statistical approaches, in contrast, model the relationships between the historical and the future power data through processing strategies for time sequences. These methods mainly include time series-oriented approaches such as autoregressive (AR) and autoregressive integrated moving average (ARIMA) models, GARCH models, among others. Experimental results associated with these methods report a prediction accuracy typically higher compared to physical methods. Finally, soft computing algorithms characterize the relationship between the inputs and the outputs, using artificial intelligence strategies such as Genetic Algorithms, Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forest (RF), and K-Nearest Neighbors (K-NN), among others [5]. Research on this topic has reported a better performance for these techniques compared to the classical physical and statistical time series models in the case of short-term wind power predictions [6].

Although interesting results have been achieved through the previously described methods, they are strongly dependent on pre-treatment strategies (Phase Space Decomposition [7], Empirical Mode Decomposition [8], etc.), where the hyper-parameters selection required during the decomposition could be a complex process leading to poor adaptability. In addition, long short-term dependencies of the time sequences are not fully

978-1-6654-8897-6/22/$31.00 ©2022 IEEE

studied. In order to overcome these limitations, deep learning (DL) strategies have been implemented in this field. Specifically, Recurrent Neural Networks (RNN) have been proposed to model the relationships among samples of the time sequences. In this context, gated RNNs (Long Short Term Memory -LSTM-, in particular) include memory units that allow controlling the flow of the relevant information associated with the relation among samples. Based on this development, several researchers have reported the implementation of hybrid methods that take advantage of this algorithm. Although most of them have reported interesting results at the local level, they are getting harder to extrapolate, do not incorporate long-term dependencies models and, report a high complexity.

In a first approach, Dong *et al*. presented the implementation of LSTM networks to forecast wind power up to 48 hours ahead, showing the increase of the error rates for each prediction step [9]. LSTM networks show a significantly better short-term forecasting performance than other classical techniques. However, forecast errors are accumulated when steps ahead are increasing due to the inherent recursive forecasting mechanism. Alternatively, Xu and Xia introduced an adaptive LSTM network, where a genetic algorithm carried out the optimization process of the hyper-parameters for the predictive model [10]. In order to reduce the prediction errors, Han *et al.* evaluated a Variational Mode Decomposition-Long Short-Term Memory (VMD-LSTM) prediction approach [11], where the wind power data were first decomposed into three constituent modes, named the long-term, the fluctuation, and the random components. Finally, LSTM networks predict each component up to 3 steps ahead. It is important to note that both long and short-term components are modeled using the same DL-based approach in this paper. Wilms *et al.*, on the other hand, presented a convolutional LSTM strategy to model spatial and temporal dependencies to analyze wind speed and wind direction data from neighboring points [12]. In the same line, a recent research developed by Ko *et al.* involved the analysis of bidirectional LSTM networks for wind power prediction [13].

Considering the successful results of LSTM networks to model long short-term dependencies and their difficulties in characterizing long-term relationships, this paper presents a hybrid approach to predict short-term wind power. The first analysis allows taking advantage of deterministic pattern recognition, identifying the long-term dynamics and the changes with a fixed and known periodicity of the wind power time series. This process provides a better understanding of data dynamics and reduces model complexity by extracting an irregular component. The second stage involves a month segmentation to better model the inner behavior, where the resulting time series are then processed by LSTM networks. This final forecasting approach allows to model time-series non-linearities, considering the high levels of abstractions provided by the associated deep structure. In this way, the ensemble strategy described in this work allows using individual advantageous effects from statistical and computational perspectives to improve the reported results in wind power forecasting. Based on real hourly wind power data from the Spanish electricity market, the effectiveness and accuracy of the method are verified. Results show an improvement in the accuracy rates for the strategy described in this work.

The following section describes the conceptual background of the forecasting method. Likewise, Section 3 presents the details of the strategy proposal of this work and its methodology. Section 4 reports the experimental design and the numerical validation, and Section 5 concludes and outlines future directions.

II. Overall Forecasting Framework

This section presents a detailed description of the different parts of the proposed methodology. First, the details associated with the processing and modeling of long-term deterministic patterns are described. Then, the long-short term dependencies are represented through a DL-based approach, namely LSTM RNNs. The final results involve the ensemble analysis of both approaches.

A. Statistical Analysis Methodology

Firstly, based on the assumption that wind power time series data are usually non-normal, a set of transforms will be studied, following the recently proposed approach developed for a long term scenarios generation in [14], to provide time series with a constant marginal variance, and normal distribution. The most popular transformations are defined in Table I.

TABLE I: NORMALIZATION TRANSFORMS

Box Cox	**Yeo Johnson**
$g(x;\lambda) = 1_{\lambda\neq 0}\frac{x^\lambda-1}{\lambda} + 1_{\lambda=0}\log(x)$	$g(x;\lambda) = 1_{\lambda\neq 0, x\geq 0}\frac{(x+1)^\lambda - 1}{\lambda} + 1_{\lambda=0,x\geq 0}\log(x+1) + 1_{\lambda\neq 2,x<0}\frac{(1-x)^{2-\lambda}-1}{\lambda-2} + 1_{\lambda=2,x<0} -\log(1-x)$
Ordered Quantile (OQ)	**Arcsinh**
$g(x) = \Phi^{-1}\left(\frac{rank(x)-\frac{1}{2}}{length(x)}\right)$	$g(x) = \log(\mathrm{x} + \sqrt{x^2+1})$

Where x refers to the original data, the parameter λ is adjusted in both cases via maximum likelihood, Φ represents the standard normal of the cumulative distribution function, and *rank*(*) and *length*(*) are the observation's rank, and the number of samples, respectively.

Subsequently, the resulting time series is decomposed into key components to model their deterministic patterns and characteristic dynamics. As such, the Seasonal-Trend decomposition using Loess (STL) is carried out to obtain the seasonal and remainder sequences of the wind power time series [15]. Firstly, taking into account an additive model, the hourly wind power time series X_t is represented as the addition of three components $X_t = T_t + S_t + R_t$. with T_t, S_t, and R_t denoting the trend, the seasonality, and the remaining components, respectively. STL is commonly defined as a filtering method to decompose time series, which applies a series of smoothing operations by means of a locally weighted regression. Then, during the weighted polynomial regression fitting, the weights decrease based on distance values from the closest neighbor

978-1-6654-8897-6/22/$31.00 ©2022 IEEE

[16]. As a result, the estimation of the smoothed time series $\widehat{X}_\iota$ is characterized by:

$$\widehat{X}_\iota = \sum_{j=0}^{d} \beta_{i,j} t_i^j \qquad j = 1, \ldots, n, \tag{1}$$

where $\beta_{i,j}$ is the $d+1$ dimensional least squares estimation for the weighted regression, t_i^j is the $d+1$ dimensional array of the period time, i is the amount of time lags up to the maximum, represented by the smoothing parameter n and d denotes the corresponding polynomial degree. Based on this configuration, the time series is recurrently fitted until the seasonality and trend components are stable. This stage involves a moving averages strategy in conjunction with the Loess Smoothing method. Finally, each component is extracted from the time series, according to:

$$T_t = T_{t+1}^{(k+1)}, S_t = S_t^{(k+1)}, R_t = X_t - T_t - S_t \tag{2}$$

with k being the number of iterations in the procedure. The most relevant advantage of this approach, in comparison with other decomposition strategies, is its strong resilience to outliers present in the time series, providing robust decomposed sub-series. In addition, STL can manage multi-seasonal time series with any seasonal frequency greater than one, not limited to only monthly or quarterly frequencies, such in alternative approaches [17]. A detailed explanation of this algorithm can be found in [15]. The resulting remainder time series R_t are then segmented for each month of the year and correspondingly analyzed by LSTM RNNs.

B. Long Short Term Memory (LSTM) Recurrent Neural Networks (RNN)

LSTM RNNs have been proposed to obtain a long-term dependencies model and the optimal time lag required during the time series analysis. In comparison with classical RNN, the memory block has been modified to control the flow of information. As such, a series of multiplicative gated units have been included inside the block. These gates can learn to allow or prevent certain data propagation, solving the classical vanishing/exploding error associated with the traditional RNNs. Figure 1 shows an LSTM memory unit and its corresponding internal structure.

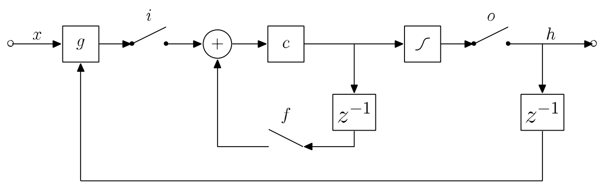

Fig. 1: Internal Structure of the LSTM block

As it can be seen, the LSTM architecture contains three gate structures, the input gate i, the forget gate f, and the output gate o. As such, i allows the information to be stored in each memory cell without any perturbation, o protects other units to be perturbed by irrelevant information, and f memory unit allows to forget irrelevant information. The corresponding activation functions for each gate are defined as:

$$i_t = \sigma(W_i . x_t + U_i . h_{t-1}), \quad f_t = \sigma(W_f . x_t + U_f . h_{t-1})$$
$$o_t = \sigma(W_o . x_t + U_o . h_{t-1}), \tag{3}$$

for each moment t, and a given input x_t, the hidden state h_t is calculated and updated by:

$$g_t = tanh(W_g . x_t + U_g . h_{t-1}), \quad c_t = c_{t-1} . f + g . i,$$
$$h_t = \tanh(c_t) . o_t, \tag{4}$$

where W and U are weight matrixes, tanh and σ are activation functions, and $t-1$ represents the former moment. New approaches reported in the literature have proposed to stack several LSTM units, achieving more accurate and efficient results [18], [19].

III. Proposed Hybrid Forecasting Approach

Based on the framework described in the previous section, the dynamical model proposed in this paper involves a combination of statistical decomposition and segmentation techniques and a computational architecture centered on stacked RNNs for each monthly segmented sequence to forecast time series with multiple seasonal patterns. In this context, our proposal can be divided into four different stages:

1. *Long Term Dependencies Extraction:* Taking into account the non-normality characteristic of real wind power time series. The first pre-processing stage involves the normalization of the skewed data. As such, four transformations formulated in Table 1, are applied to the data, and the transformation with the best performance is finally selected for the analysis. Based on the selected normalization transform, the resulting time series is represented as a combination of trend, seasonal, and remainder components by means of the implementation of the STL transform described in Section II-A. During the decomposition process, two seasonality frequencies are considered for this wind power time series: annual (8760 hours) and daily (24 hours). In addition, an additive decomposition for the wind power time series is consequently explored. After the decomposition process is completed, the remainder component is extracted to continue the analysis.

2. *Time Series Segmentation:* Pre-processed results associated with the remainder values, obtained from the real-time series when seasonal and trend components are subtracted, are monthly segmented to improve the accuracy of the wind power time series prediction for each monthly evaluation.

3. *Stacked LSTM RNN:* A stacking architecture to train an LSTM RNN is proposed to extract the remaining dependencies for each month. In this stage, the same network configuration is preserved, weights will change based on the time series training process carried out for each month.

4. *De-normalization and Trend and Seasonal Effect Inclusion:* Final predictions are obtained after an inverse transformation, associated with the de-normalization process and the addition of the trend and seasonal effect to the wind power time series for each respective month.

978-1-6654-8897-6/22/$31.00 ©2022 IEEE

IV. EXPERIMENTAL RESULTS

To carry out the validation, firstly, the wind power data and the evaluation measures to be used in the study are defined and analyzed. Then, the performance of the proposed forecasting model is validated on real wind power time series for 1, 2, and 3 steps ahead. All the experimental analyses are performed in the R (v. 3.6.2) and Python (v. 3.7.4) environment on a 3.5 GHz PC with process Intel Xeon E3-1246 and 32 GB RAM.

A. Evaluation Metrics

The performance evaluation of the proposed approach is performed by comparing the Root Mean Square Error (RMSE), the Mean Absolute Percentage Error (MAPE), and the Mean Absolute Error (MAE), defined by:

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(x_n - \hat{x}_n)^2} \tag{5}$$

$$MAPE = \frac{1}{N}\sum_{n=1}^{N}\frac{|x_n - \hat{x}_n|}{x_n} \times 100\% \tag{6}$$

$$MAE = \frac{1}{N}\sum_{n=1}^{N}|x_n - \hat{x}_n|, \tag{7}$$

where N is the number of predicted samples, x_n and $\hat{x}_n$ are the real value and predicted values, respectively. RMSE has been widely used in meteorological variables forecasting because it satisfies the objectivity and symmetry condition. The MAE metric, on the other hand, weighs all values equally and does not give extreme forecasting events any extra weight. Finally, MAPE has been proposed as the most popular unit-free accuracy metric for predicting problems, and several textbooks, M-competition, and the prior literature motivate its use [20].

B. Input Data

Experiments involved in this study use real hourly wind power data from 2008-2019 from the Spanish electricity market. Table II presents the sample descriptive statistics corresponding to the real-time series analyzed in this work.

TABLE II: DESCRIPTIVE STATISTICS OF THE INITIAL TIME SERIES

Min	**Max**	**Mean**	**SD**	**Skewness**	**Kurtosis**
0.0093	0.7704	0.2469	0.1394	0.7419	-0.0290

Based on the skewness calculation (positive), the wind power time series analyzed are not normal, showing a right-skewed distribution. Kurtosis value (less than three) indicates a platykurtic distribution (flatter than a normal distribution). A normalization transform is implemented on the time series based on this dynamics analysis.

C. Statistical Preprocessing Results

To normalize data and stabilize variance in the real-time series presented in this work, four widely implemented transforms are explored (Table I). To select the best transformation for the analyzed time series, the Pearson P statistic (divided by its degrees of freedom) is calculated. This normalization measure is selected because it allows to compare between transformations as an absolute distance of the departure of normality, and it is a relatively interpretable goodness-of-fit measure. In this case, if the data is close to a normal distribution, this measure will be close to 1. Results found for this real-time series for each transformed and the non-transformed (NT) data are summarized in Table III. Results show a better-normalized distribution (closer to 1) for the transformed data using the Ordered Quantile approach. Statistical metrics are shown in Table IV.

TABLE III: PEARSON STATISTICS

NT	**Box Cox**	**Yeo Johnson**	**OQ**	**ArcSinh**
23.955	3.767	8.724	**1.125**	21.368

TABLE IV: DESCRIPTIVE STATISTICS OF THE NORMALIZED TIME SERIES

Min	**Max**	**Mean**	**SD**	**Skewness**	**Kurtosis**
-4.420	4.420	0.000	1	0.000	-0.001

This transformed data is then decomposed to remove its trend and seasonal components. As such, the STL decomposition approach (Section II-A) is applied to the data.

D. Time Series Segmentation Results

Remainder values extracted from the previous decomposition stage are the inputs of the RNN-based forecasting method described in this work. As such, in order to fine tune the RNN implemented to predict the wind power values, a monthly characterization of the time series is proposed. In this case, the remainder component is monthly segmented by generating 12 training/testing datasets that will customize the RNN weights for each month of the year. It is important to note that the RNN architecture will be preserved for every month, just weights will change to characterize the dynamics of every month. Twelve datasets are consequently structured with 1, 2, and 3 forecasting steps ahead for validation. Correspondingly, the final datasets are divided into training and testing data with two-thirds and one-third of the complete datasets.

E. Stacked RNN Prediction and Final Forecasting Results

The data for the sequence prediction problem studied in this paper needs to be scaled [0 to 1] when training neural networks, such as LSTM RNN. The main reason lies in the fact that when a network is fit on non-scaled data, the learning and convergence of the RNN can slow down and, in some cases, prevent the network from effectively learning the time series dependencies. As such, an iterative process with the whole data was first implemented to find the architecture with the best performance for an LSTM RNN. Two stacked networks were developed to model the remaining values of the wind power time series. During the RNNs training, the number of hidden neurons for each network was determined by cross-validation to be in the range of 16, 32, or 64 neurons. As a result, the number of hidden neurons that provide the best performance on

978-1-6654-8897-6/22/$31.00 ©2022 IEEE

validation data was finally selected (32 for the first RNN and 16 for the second one). Hidden neurons are activated by the function ReLU (Regular Linear Unit), and recurrent output neuron uses a hard sigmoid. Each neural network is trained using the Adam optimization algorithm in conjunction with a cross-validation-based early stopping strategy to prevent over-fitting. In addition, the Mean Square Error (MSE) is the loss function used to be minimized. Considering this configuration, the training process with monthly segmented time series is carried out for each month.

Results are correspondingly re-scaled. Subsequently, the seasonal and trend components, extracted in Section IV-C, are used to estimate future values through a naive seasonal method and a combination of STL and Exponential Smoothing (ETS), respectively. This selection is widely implemented based on the successful reported results at analyzing seasonal time series with non-linear trend-cycles [21]. These values are incorporated, and as the last step, inverse Ordered-Quantile transform is applied to calculate the final prediction values. RMSE, MAPE and MAE errors are calculated during the validation process for each month for 1, 2 and 3 steps ahead (See Figures 2, 3 and 4, respectively). Based on the error values shown in the figures, it is possible to see that February, April, August, and November are the months with more difference errors for the three evaluation measures defined in this work, for the one-step ahead case. In contrast, July is the month that preserves a small error for the three evaluation measures. This distribution changes for two and three steps ahead, where March and December, and April and October are the months with more variability, respectively. It is important to highlight the error measures variability for each month, showing that the monthly inner dynamics are likely to be better captured by the proposed approach for certain specific months. In conclusion, small errors are found for one-step ahead (Average $MAPE = 5.1\%$), and as it is expected, differences increase for two and three steps ahead (Average $MAPEs = 8.2\%$ and 12.9%, respectively); which is an inherent characteristic for the LSTM-based approaches, when several steps ahead are considered.

To investigate the effectiveness of the proposed hybrid model, it is compared with a stacked LSTM that will analyze the raw data associated with wind power in Spain. Because the previous LSTM architecture was selected based on the complete time series, the same configuration is implemented with two networks and 32 and 16 neurons, respectively. Results are summarized in Figure 5. Based on the experimental results, it can be observed that the proposed hybrid forecasting method shows superior performance compared with traditional stacked LSTM RNNs (5.101 % to 11.648 % for one-step-ahead MAPE errors), which deeply illustrates that the proposed methodology is a relevant DL-based alternative for short-term wind power forecasting.

V. CONCLUSIONS

Considering the increasing influence of wind energy, wind power forecasting methodologies have become crucial for the safety management and operation of power systems. This study proposes an ensemble methodology based on statistical pre-processing methodologies and DL-based algorithms to achieve highly accurate short-term forecasting results. A combination of OQ Normalization, STL decomposition, and LSTM RNN has been implemented as an alternative solution. This last algorithm has been fine-tuned with different training datasets associated with the segmentation for each month of the year. To validate this approach, it has been compared with a DL-based traditional strategy: the LSTM RNN, which has been recently used to forecast these characteristic time series. Then, a stacked LSTM with two RNNs is implemented, as in the last step of our proposed approach. For this process, real hourly wind power data from the Spanish electricity market has been analyzed to forecast one, two, and three steps ahead. Experiments results indicate that the proposed methodology can achieve more accurate wind power forecasting results than the single model.

REFERENCES

[1] Y. Wang, R. Zou, F. Liu, L. Zhang, and Q. Liu, "A review of wind speed and wind power forecasting with deep neural networks," Applied Energy, vol. 304, p. 117766, 2021.

[2] Y. Zhao, L. Ye, P. Pinson, Y. Tang, and P. Lu, "Correlation-constrained and sparsity-controlled vector autoregressive model for spatio-temporal wind power forecasting," IEEE Transactions on Power Systems, vol. 33, no. 5, pp. 5029–5040, 2018.

[3] A. Ahmed and M. Khalid, "A review on the selected applications of forecasting models in renewable power systems," Renewable and Sustainable Energy Reviews, vol. 100, pp. 9–21, 2019.

[4] S.-l. Feng, W.-s. Wang, C. Liu, and H.-z. Dai, "Study on the physical approach to wind power prediction [j]," Proceedings of the CSEE, vol. 30, no. 2, pp. 1–6, 2010.

[5] C.-M. Huang, C.-J. Kuo, and Y.-C. Huang, "Short-term wind power forecasting and uncertainty analysis using a hybrid intelligent method," IET Renewable Power Generation, vol. 11, no. 5, pp. 678–687, 2017.

[6] P. Du, J. Wang, W. Yang, and T. Niu, "A novel hybrid model for shortterm wind power forecasting," Applied Soft Computing, vol. 80, pp. 93–106, 2019.

[7] L. Han, C. E. Romero, and Z. Yao, "Wind power forecasting based on principle component phase space reconstruction," Renewable Energy, vol. 81, pp. 737–744, 2015.

[8] W. Sun and Y. Wang, "Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network," Energy conversion and management, vol. 157, pp. 1–12, 2018.

[9] D. Dong, Z. Sheng, and T. Yang, "Wind power prediction based on recurrent neural network with long short-term memory units," in 2018 International Conference on Renewable Energy and Power Engineering (REPE). IEEE, 2018, pp. 34–38.

[10] G. Xu and L. Xia, "Short-term prediction of wind power based on adaptive lstm," in 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2). IEEE, 2018, pp. 1–5.

[11] L. Han, R. Zhang, X. Wang, A. Bao, and H. Jing, "Multi-step wind power forecast based on vmd-lstm," IET Renewable Power Generation, vol. 13, no. 10, pp. 1690–1700, 2019.

[12] H. Wilms, M. Cupelli, A. Monti, and T. Gross, "Exploiting spatiotemporal dependencies for rnn-based wind power forecasts," in 2019 IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia). IEEE, 2019, pp. 921–926.

[13] M.-S. Ko, K. Lee, J.-K. Kim, C. W. Hong, Z. Y. Dong, and K. Hur, "Deep concatenated residual network with bidirectional lstm for onehour-ahead wind power forecasting," IEEE Transactions on Sustainable Energy, vol. 12, no. 2, pp. 1321–1335, 2020.

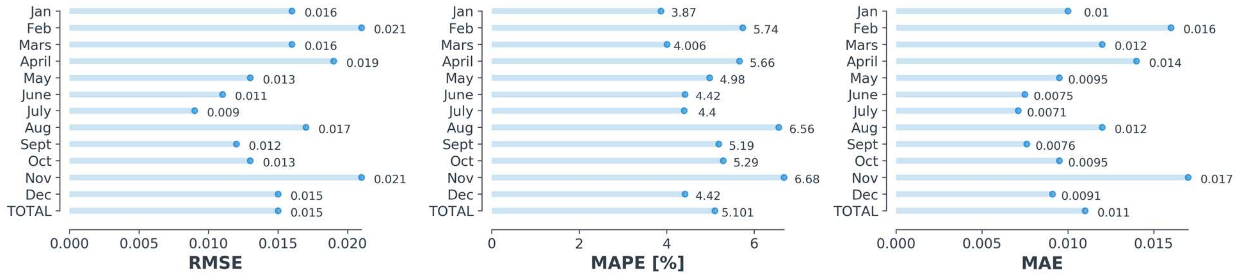

Fig. 2: One-step ahead forecasting errors

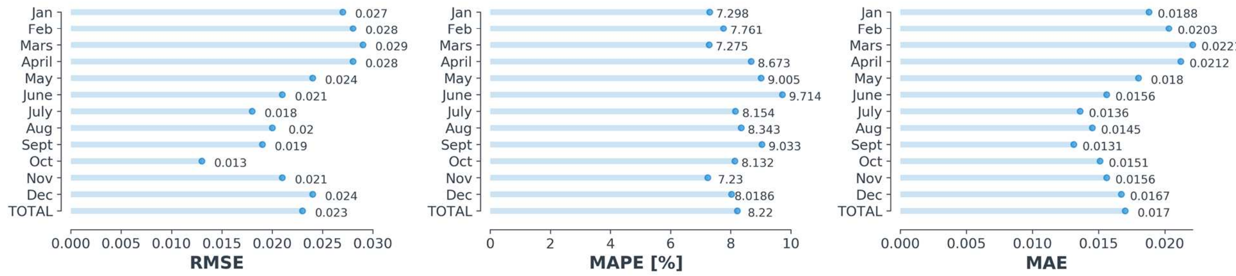

Fig. 3: Two-step ahead forecasting errors

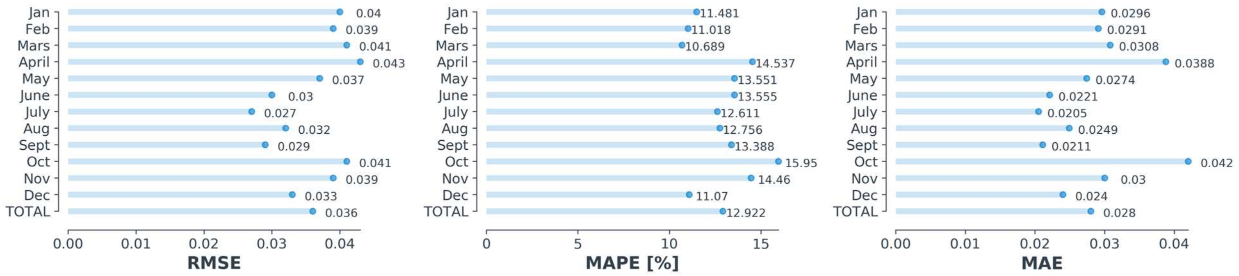

Fig. 4: Three-step ahead forecasting errors

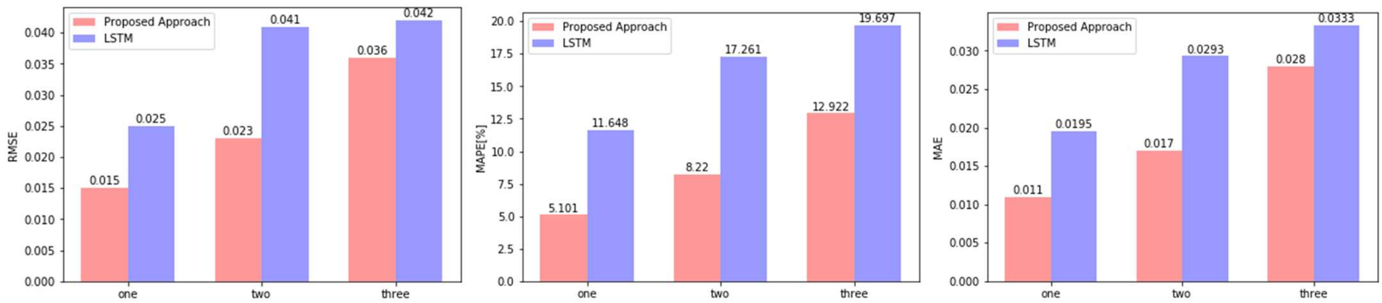

Fig. 5: Error comparison with a stacked LSTM approach

[14] G. Marulanda, A. Bello, J. Cifuentes, and J. Reneses, “Wind power longterm scenario generation considering spatial-temporal dependencies in coupled electricity markets,” Energies, vol. 13, no. 13, p. 3427, 2020.

[15] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, “Stl: A seasonal-trend decomposition,” Journal of official statistics, vol. 6, no. 1, pp. 3–73, 1990.

[16] E. B. Dagum and A. Luati, “Global and local statistical properties of fixed-length nonparametric smoothers,” Statistical Methods and Applications, vol. 11, no. 3, pp. 313–333, 2002.

[17] M. Theodosiou, “Forecasting monthly and quarterly time series using stl decomposition,” International Journal of Forecasting, vol. 27, no. 4, pp. 1178–1195, 2011.

[18] Z. Cui, R. Ke, Z. Pu, and Y. Wang, “Deep bidirectional and unidirectional lstm recurrent neural network for network-wide traffic speed prediction,” arXiv preprint arXiv:1801.02143, 2018.

[19] M. Abdel-Nasser and K. Mahmoud, “Accurate photovoltaic power forecasting models using deep lstm-rnn,” Neural Computing and Applications, vol. 31, no. 7, pp. 2727–2740, 2019.

[20] L. Chhay, M. A. H. Reyad, R. Suy, M. R. Islam, and M. M. Mian, “Municipal solid waste generation in china: influencing factor analysis and multi-model forecasting,” Journal of Material Cycles and Waste Management, vol. 20, no. 3, pp. 1761–1770, 2018

[21] R. J. Hyndman and G. Athanasopoulos, Forecasting: principles and practice.

978-1-6654-8897-6/22/$31.00 ©2022 IEEE