Short-Term Wind Energy Forecasting with Temporally Dependent Neural Network Models

Rui Li
SAS Institute Inc.
500 SAS Campus Dr.
Cary, NC 27519
rui.li@sas.com

Pu Wang
SAS Institute Inc.
121 W Trade Street.
Charlotte, NC 28202
pu.wang@sas.com

Jingrui Xie
SAS Institute Inc.
500 SAS Campus Dr.
Cary, NC 27519
rain.xie@sas.com

Alex Chien
SAS Institute Inc.
500 SAS Campus Dr.
Cary, NC 27519
alex.chien@sas.com

Mustafa Kabul
SAS Institute Inc.
500 SAS Campus Dr.
Cary, NC 27519
mustafa.kabul@sas.com

ABSTRACT

As the penetration of renewable energy into the electrical grid is increasing worldwide, accurate forecasting of renewable energy generation is essential not only for grid operation and reliability, but also for energy trading and long-term planning. In this paper, we focus on short-term wind energy forecasting. The inherent variability and unpredictability of wind energy imposes great challenges upon many models. Conventional time series models, such as ARIMAX, often fail to capture nonlinear patterns in energy output, and a feedforward artificial neural network doesn't take temporal dependency into account. In this paper, we apply state-of-art autoregressive artificial neural network (AR-ANN) models and recurrent neural network (RNN) models to wind energy forecasting. By capturing both the sequential pattern of energy output and the complex relationship between weather predictors and power generation, we can achieve better forecasting accuracy. These temporally dependent neural network structures can also be easily extended to model other nonlinear time series and temporal data.

KEYWORDS

recurrent neural network, autoregressive neural network, wind energy forecasting, renewable energy

1 INTRODUCTION

Among the ever-increasing share of renewable energy in the global energy portfolio, the deployment of wind energy–generating capacities has been most rapid and consistent worldwide [18, 20]. In contrast to conventional energy generation, wind energy generation is largely uncontrollable and highly variable due to the volatile nature of its energy source—wind. As a result, wind energy generation is often balanced and backed up by ancillary generators, which creates operational challenges and increases costs [8]. It becomes obvious that accurate wind energy forecasting, especially short-term forecasting, is crucial for utilities in order to increase the integration of wind power into the electrical grid while maintaining its reliability at a reasonable cost.

Wind energy forecasting has also received great attention from academia in recent years [10, 13] because it represents a general time series problem that has two important characteristics: (1) The data often contains covariates that allow for a regression-based analysis. In wind energy forecasting, these covariates include future wind speed and wind direction that are predicted by weather data providers. The fact that these covariates are predicted values instead of measured values and contain prediction errors increases the challenge of modeling the complex relationship between them and the wind energy output. (2) Wind energy generation is also autocorrelated in time and shows a time-dependent pattern [17]. Assuming independence in regression models is not realistic and often results in poor forecasting performance.

In recent years, numerous research studies have been conducted for wind energy forecasting and can be generally summarized into two categories. The first one is conventional time series models, such as autoregressive integrated moving average (ARIMA) model, and its variant ARIMAX, which allows the inclusion of covariates [26]. Although ARIMAX can handle both the regression on covariates and temporal dependency, it assumes that the response variable at the current time step has a linear relationship with the covariates, response variables and error terms at previous time steps. This assumption is often violated in wind energy forecasting, because the characteristic curve of wind energy generation with respect to wind speed is "sigmoid-shaped" as opposed to linear [21]. In addition, a number of hyperparameters must be tuned prior to modeling; these hyperparameters include the orders of autoregressive model and moving average model, as well as the degree of differencing [16]. This tuning increases the burden for forecasters as they strive to find the optimal model. Another category of models that has received a lot of attention recently are machine learning methods, which are represented by the feedforward artificial neural network (ANN) [7, 11]. ANN is flexible and capable of modeling the nonlinear and complex relationship between wind energy output and covariates such as the wind speed and wind directions. One obvious drawback of direct modeling with a feedforward artificial neural network is the lack of temporal information, which could be useful in very short-term forecasting.

In this article, we propose two variations of models that are based on artificial neural networks to address the limitations of the conventional time series models and machine learning methods. Our focus is on short-term forecasting of wind energy where it has the most value in electrical grid operation. More specifically, our strategies and contributions are:

(1) We examine the residual temporal patterns after modeling with a feedforward neural network, and we show a strong temporal-dependency in the wind energy output.
(2) We propose the autoregressive neural network model, into which lagged wind energy outputs are suitably incorporated in order to handle autocorrelations and improve forecasting accuracy over conventional ARIMAX and feedforward neural network methods.
(3) In addition, we propose a recurrent neural network method, which takes both the current covariates and the historical sequence of wind energy output into consideration and improves forecasting accuracy over conventional methods.
(4) We compare the forecasting performance in both one-step-ahead and multi-step-ahead forecasting scenarios. We use both recursive one-step-ahead forecasting and sequence-to-sequence translation in a recurrent neural network to achieve multi-step-ahead forecast. Our models outperform conventional models in multi-step-ahead forecasting, which is crucial in wind energy forecasting practice.
(5) We also provide our suggestions of the advantages and disadvantages for each method in multi-step-ahead forecasting and a detailed analysis of the phenomenon of "accumulation of error" in multi-step-ahead forecasting.

The remainder of the paper is organized as follows: We discuss about the related work in section 2. We describe the wind energy data in section 3. In section 4, we introduce the two models proposed in this research. In section 5, we demonstrate our modeling strategies and detailed analysis on model performance by comparing them to the benchmark models. In section 6, we discuss the implications of our approach and its impact on model choice, model tuning, and short-term wind energy forecasting.

2 RELATED WORK

Global Energy Forecasting Competition 2012 (GEFcom2012) has attracted hundreds of contestants worldwide and motivated the development of various methods in short-term wind energy forecasting during and after the competition [9]. These methods include, but not limited to: gradient boosting models [22], artificial neural network models [12, 25], Gaussian Process models [6, 12] and k-nearest neighbors [15].

The reported methods are predominantly regression and machine learning models rather than time series models, partially because the original testing period in the competition is discontinuous in time and give time series models a disadvantage.

Our work focuses on the incorporation of time series components into existing machine learning models and evaluate how it will affect the performance in one-step-ahead and multi-step-ahead forecasting scenarios. Regular neural network model that is frequently used by contestants, serves as one of the benchmark in our research.

3 DATA DESCRIPTION

The wind energy data in this article is from the wind track of the Global Energy Forecasting Competition 2012 (GEFcom2012) [9]. It contains historical wind energy measurements at seven wind farms and meteorological forecasts of wind components at the level of wind farms. Zonal (u) and meridional (v) components of surface wind at 10m above ground level are included in the original data, so we can derive the wind speed as $ws = \sqrt{u^2 + v^2}$. The wind energy generation is scaled to the nominal capacities of each farm in the original competition data, which allows for a scale-free comparison across different wind facilities. The scatter plot in figure 1 shows the relationship between wind speed and wind energy generation in farm 1. The plot demonstrates that for a particular wind speed, wind energy output can vary within a wide range. This variation can come from other covariates, from unobserved components, or from the varying quality of wind speed forecast itself. In addition, the underlying power curve appears to be "sigmoid-shaped", indicating that the assumptions of linear regression based methods are violated.

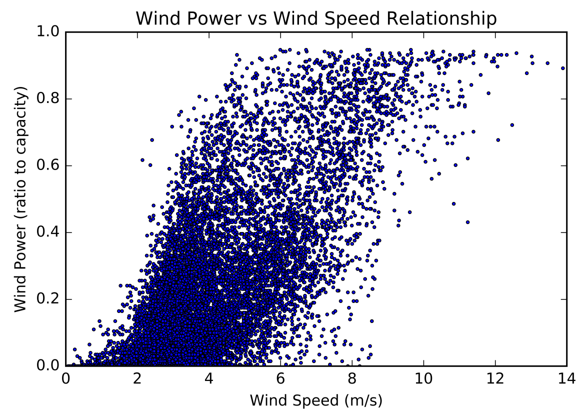

Figure 1: Relationship between wind speed and wind energy generation.

4 PROPOSED MODELS

In this section we propose two methods that are based on neural networks, with a focus on modeling two aspects of wind energy generation: (1) the complex relationship between wind energy and weather covariates such as wind speed, wind direction, and so on, and (2) the temporal dependency in wind energy generation. We examine both models and compare them to the conventional feedforward neural network model and ARIMAX in 1-hour-ahead forecasting and 24-hour-ahead forecasting in Section 5.

4.1 Autoregressive Artificial Neural Network (AR-ANN)

Artificial neural network (ANN) is a popular machine learning approach to handling a variety of problems, including classification, anomaly detection, regression, forecasting, and so on [4]. ANN does not assume a linear relationship between the covariates and the response variable, and it has great flexibility in terms of approximating any underlying continuous functions because of its universal approximation property with mild assumptions on the activation functions [2]. In order to account for temporal dependency in an autoregressive artificial neural network [24], we include the wind energy output at previous time steps as covariates in the model. To forecast the wind energy y_t at time t, the model formulation is

$$\hat{y}_t = a_0 + \sum_{j=1}^{k} \beta_j \max[0, b_{j0} + \sum_{i=1}^{p} w_{ij} x_{it} + \sum_{l=1}^{d} \alpha_{lj} y_{(t-l)}], \quad (1)$$

where y_t is the wind energy output at time t, x_{it} is the ith covariate at time t, k is the number of neurons in the hidden layer, p is the number of covariates collected from weather forecast vendors or engineered by users, and d is the number of lagged wind energy output to include in the model. The weights to be trained in the neural network training are $\{a_0, \beta_j, b_{j0}, w_{ij}, \alpha_{lj}\}$, where $j \in \{1, 2, \cdots, k\}$, $i \in \{1, 2, \cdots, p\}$, and $l \in \{1, 2, \cdots, d\}$. As indicated in equation (1), we use rectified linear units (ReLU) as an activation function in this analysis. For AR-ANN and regular ANN models, we use one hidden layer with 60 neurons and the dropout rate is set at 0.35.

For multi-step-ahead forecasting with an autoregressive artificial neural network, we perform a one-step-ahead forecast recursively. For example, when we try to forecast wind energy output at both time t and time $t+1$, we obtain $\hat{y}_t$ from equation (1), and plug $\hat{y}_t$ into the right-hand side of equation (1) to obtain $\hat{y}_{t+1}$, because the actual y_t is not observed at the time when we generate the forecast.

4.2 Recurrent Neural Network (RNN)

Recurrent neural network is a specialized neural network structure, whose cyclic connection is designed for modeling the sequence of events [5]. It has been used in speech recognition, text mining, translation, and time series forecasting [3, 14]. To forecast wind energy output y_t at time t, we adopt the structure shown in Figure 2, so that we can combine the sequential information of wind energy output with the information from covariates such as forecasted wind speed and direction at time t. In this work, we have two RNN hidden layers at each time step with 30 and 20 hidden neurons in each layer. Only one layer is drawn in Figure 2 for demonstration purposes. The feedforward layer is constructed the same way as in ANN or AR-ANN.

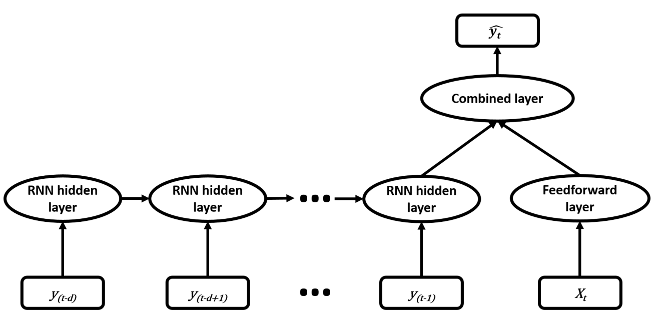

Figure 2: Recurrent neural network structure for a one-step-ahead forecast. The RNN hidden layer uses the long short term memory (LSTM) structure, y_t is the wind energy output at time t, and X_t are the covariates for time t.

Recurrent neural network has also been applied to sequence-to-sequence translation [1, 23], where the "encoder-decoder" structure is used to translate an input sequence into an output sequence with a different length. We adapt this structure for multi-step-ahead forecasting, where the output sequence length is the forecasting horizon. The adapted structure is shown in Figure 3. The encoding and decoding layers each has 15 hidden neurons.

To prepare samples for a recurrent neural network, we use the sliding windows approach to process the time series into subsequences and we stack the subsequences into three-dimensional arrays whose dimensions are $observations \times timesteps \times features$. The number of features for the response variable series is 1 and the number of features for the covariates series is the number of covariates p. This preprocessing converts long time series into independent subsequences without losing temporal information within each subsequences. It allows easy implementation of parallelization, in contrast to conventional time series models such as ARIMAX.

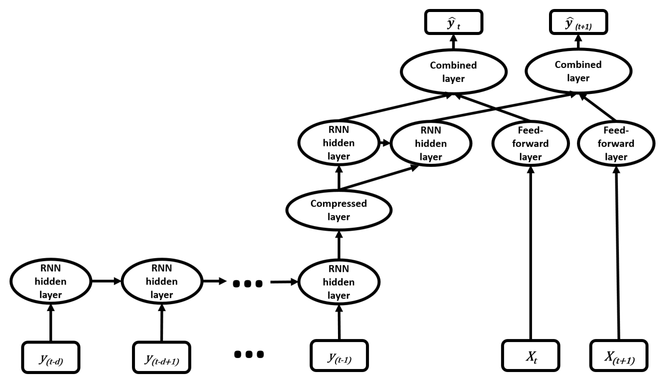

Figure 3: Recurrent neural network structure for a multi-step-ahead forecast (two-step-ahead is used here for an example). The compressed layer is encoded by a series of RNN hidden layers from the input, and then decoded by another series of RNN hidden layers to produce an arbitrary output sequence length.

5 RESULTS

We test the performance of our models on the wind energy data described in section 3 and compare them with two benchmark models: (1) a feedforward artificial neural network that includes all the collected covariates, and (2) an ARIMAX model that includes the collected covariates as exogenous variables.

Prior to applying our proposed AR-ANN or RNN, we examined the temporal pattern of the wind energy output residuals after modeling with a feedforward artificial neural network. The autocorrelation function (ACF) and partial autocorrelation function (PACF) are shown in Figure 4, a representative figure that uses the data from farm 1.

It is clear that the residuals are autocorrelated, indicating that directly applying the feedforward artificial neural network has missed important temporal information. In addition, we observe

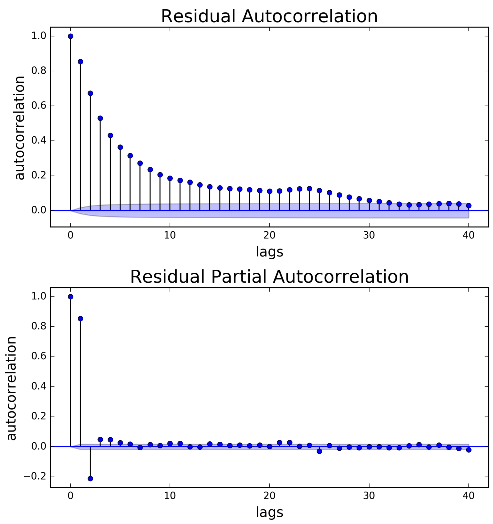

Figure 4: ACF and PACF for the residuals after modeling with a feedforward artificial neural network.

that the residual PACF has two lags of obvious partial autocorrelation and that the partial autocorrelation is significantly diminished beyond lag 2. We propose that this information is indicative of the length of historical information that we need to consider in both AR-ANN models and RNN models. To test our hypothesis, we use both AR-ANN and RNN models with different numbers of lags or history sequence lengths to make one-hour-ahead forecasts for a half-month period at the end of the dataset (2011 Dec. 15 to 2011 Dec. 31). We also check the residual ACF and PACF for the training dataset and compare them across different lags or history sequence lengths. The results are shown in Figure 5 through Figure 8.

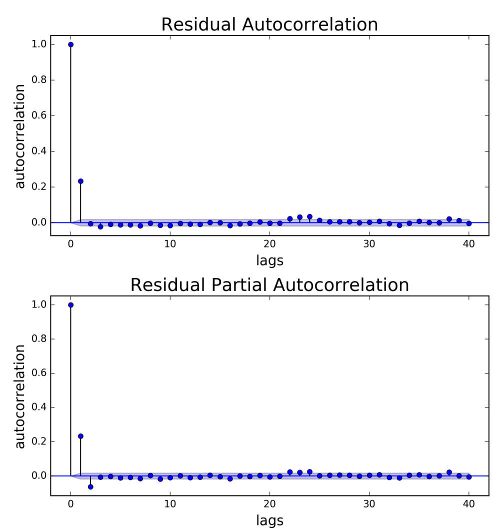

Figure 5: AR-ANN with lag1 of wind energy output used as a covariate.

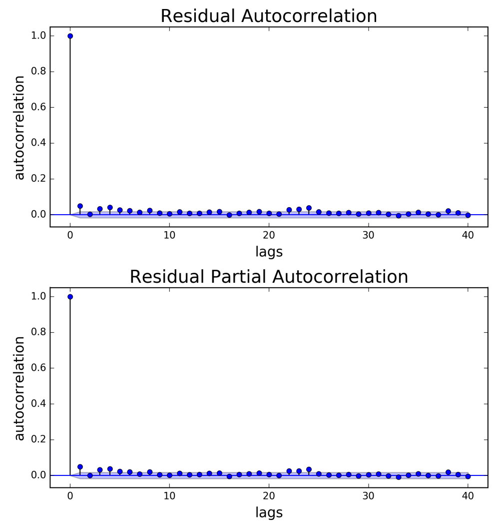

Figure 6: AR-ANN with up to lag2 of wind energy output used as covariates.

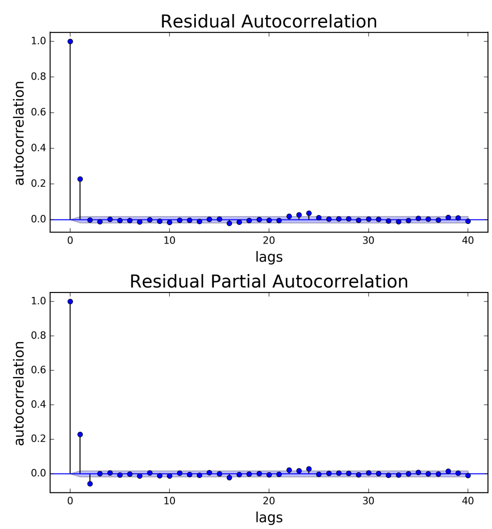

Figure 7: RNN with history sequence length = 1

Consistent with our hypothesis, the number of lags used in AR-ANN being equal to 2 or the history sequence length in RNN being equal to 2 is sufficient to remove residual autocorrelation.

We further examine whether removing residual autocorrelation helps improve the accuracy in one-step-ahead forecasting (one-hour-ahead in this example). The forecasting accuracy, measured by root mean squared error (RMSE) is shown in Figure 9.

The results show that when the number of lags in AR-ANN or the history sequence length in RNN is larger than 2, the forecasting performance reaches a plateau, which corresponds to the complete removal of residual autocorrelation. However, when the number of lags or the history sequence length is smaller, and the residual autocorrelation is not completely removed, as in Figure 5 and Figure

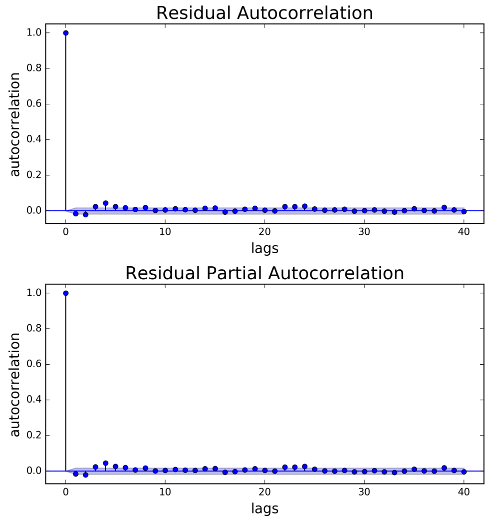

Figure 8: RNN with history sequence length = 2

7 respectively, the forecasting is less accurate and might still be improved by increasing the number of lags or history sequence length.

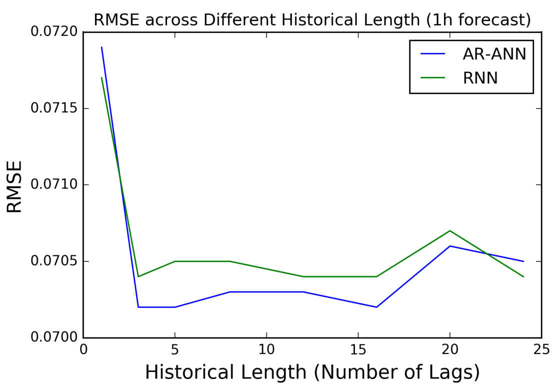

Figure 9: Average forecasting error of 7 farms across different history sequence length (number of lags). The shortest sequence length tested is 1 and longest is 24.

After we know the appropriate history length or number of lags to include in the model, we can compare the forecasting performance of our proposed AR-ANN and RNN with the feedforward artificial neural network and ARIMAX. The results are included in Table 1. Apparently, including historical information either through an AR-ANN model or a RNN model can significantly improve forecasting accuracy over conventional ANN in one-step-ahead forecasting. ARIMAX, which also depends on observations at previous time steps, achieves smaller errors compared to ANN. This observation indicates that in a one-hour-ahead wind energy forecast, historical information plays an important role in forecasting accuracy.

Table 1: One-Hour-Ahead Forecasting Errors

Models	ARIMAX	ANN	AR-ANN	RNN
RMSE	0.0723	0.1701	0.0702	0.0704

Another important, yet very challenging, area in wind energy forecasting is day-ahead forecasting, which is crucial for grid balancing and operational planning [19]. To test our models in this application, we perform multi-step-ahead (24-hour-ahead) forecasting and compare with conventional methods. ARIMAX, AR-ANN, and RNN all depend on the wind energy output at previous time steps to make a forecast for the next time step. Thus in multi-step-ahead forecasting, predicted values from one-step-ahead forecast are plugged back into the model to forecast the next step. Doing this incurs the phenomenon called the "accumulation of error", especially for models that depend heavily on previous response variable values. As shown in Table 2, except for ANN, which doesn't depend on response variables at a previous time step to make a forecast, all other models have increased forecasting errors. However, AR-ANN and RNN still show superior forecasting accuracy compared to ANN, probably because of the excellent accuracy that is achieved in the one-step-ahead forecast.

Table 2: 24-Hour-Ahead Forecasting Errors

Models	ARIMAX	ANN	AR-ANN	RNN
RMSE	0.194	0.170	0.151	0.148

Similarly, we investigate the impact of history sequence length or number of lags on a multi-step-ahead forecast in Figure 10. In contrast to one-step-ahead forecasting, the longer the history that is included in our model, the worse the performance in a multi-step-ahead forecast. This is because models that have long history sequence length rely more heavily on the response observations in previous time steps and thus are more susceptible to the "accumulation of error".

With the ability to perform sequence-to-sequence translation in RNN as described in Section 4.2, we can directly choose the output sequence length to be the forecast horizon (24 hours in this example). This will avoid doing recursive one-step ahead forecasting and thus the "accumulation of error". As shown for the RNN (seq-to-seq) case in Figure 10, different input lengths have minimum impact on forecasting accuracy when the sequence-to-sequence translation is used. Although in this 24-hour-ahead forecasting experiment, sequence-to-sequence translation doesn't show superior performance compared to AR-ANN or RNN with recursive one-step-ahead forecasting (if an appropriate history length is chosen), it does provide a viable alternative in situations when the "accumulation of errors" is severe.

6 CONCLUSION

In this article, we propose and apply two temporally dependent neural network models, AR-ANN and RNN, for the wind energy forecasting problem. Both methods are designed to capture the temporal patterns of wind energy generation, and at the same time

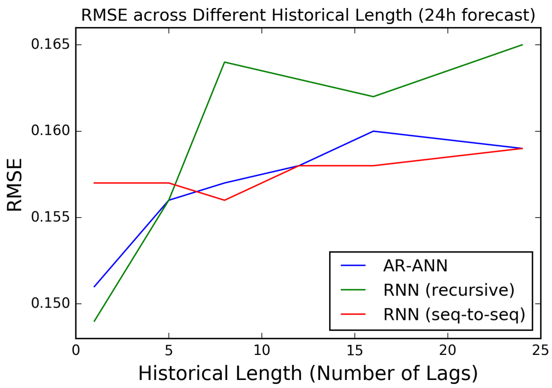

Figure 10: Forecasting errors in a multi-step-ahead forecast across different history sequence lengths (number of lags). The shortest sequence length tested is 1, and the longest is 24.

model the complex relationship between weather covariates and wind energy output. The ACF and PACF analysis suggests that residual autocorrelation is an important indicator and a potential source of error for wind energy forecasting. The PACF plot also provides additional information for tuning the necessary length of the history to be included in the temporally dependent neural network models, mirroring the common practice in autoregressive moving average (ARMA) models. Both proposed models also achieve better forecasting performance in 1-hour-ahead and 24-hour-ahead wind energy forecasting, compared to conventional ANN and ARIMAX. In addition, we also use the sequence-to-sequence translation structure in RNN for multi-step-ahead forecasting, and we suggest it is a proper alternative to avoid the "accumulation of error" phenomenon.

The scope of this paper prevents us from extensively investigating additional benefits of the proposed models. One of these additional benefits over ARIMAX is from the computational aspect, where the structures of both AR-ANN and RNN allow for easy implementation of parallelization and scale up to large datasets that contain long time series. We will explore these benefits in future studies.

REFERENCES

[1] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* (2014).

[2] Balázs Csanád Csáji. 2001. Approximation with artificial neural networks. *Faculty of Sciences, Etvs Lornd University, Hungary* 24 (2001), 48.

[3] Wim De Mulder, Steven Bethard, and Marie-Francine Moens. 2015. A survey on the application of recurrent neural networks to statistical language modeling. *Computer Speech & Language* 30, 1 (2015), 61–98.

[4] Howard B Demuth, Mark H Beale, Orlando De Jess, and Martin T Hagan. 2014. *Neural network design.* Martin Hagan.

[5] Jeffrey L Elman. 1990. Finding structure in time. *Cognitive science* 14, 2 (1990), 179–211.

[6] Shengchen Fang and Hsiao-Dong Chiang. 2017. A high-accuracy wind power forecasting model. *IEEE Transactions on Power Systems* 32, 2 (2017), 1589–1590.

[7] Aoife M Foley, Paul G Leahy, Antonino Marvuglia, and Eamon J McKeogh. 2012. Current methods and advances in forecasting of wind power generation. *Renewable Energy* 37, 1 (2012), 1–8.

[8] Hannele Holttinen, Michael Milligan, Erik Ela, Nickie Menemenlis, Jan Dobschinski, Barry Rawn, Ricardo J Bessa, Damian Flynn, Emilio Gomez-Lazaro, and Nina K Detlefsen. 2012. Methodologies to determine operating reserves due to increased wind power. *IEEE Transactions on Sustainable Energy* 3, 4 (2012), 713–723.

[9] Tao Hong, Pierre Pinson, and Shu Fan. 2014. Global energy forecasting competition 2012. (2014).

[10] Jaesung Jung and Robert P Broadwater. 2014. Current status and future advances for wind speed and power forecasting. *Renewable and Sustainable Energy Reviews* 31 (2014), 762–777.

[11] GN Kariniotakis, GS Stavrakakis, and EF Nogaret. 1996. Wind power forecasting using advanced neural networks models. *IEEE transactions on Energy conversion* 11, 4 (1996), 762–767.

[12] Duehee Lee and Ross Baldick. 2014. Short-term wind power ensemble prediction based on Gaussian processes and neural networks. *IEEE Transactions on Smart Grid* 5, 1 (2014), 501–510.

[13] Ma Lei, Luan Shiyan, Jiang Chuanwen, Liu Hongling, and Zhang Yan. 2009. A review on the forecasting of wind speed and generated power. *Renewable and Sustainable Energy Reviews* 13, 4 (2009), 915–920.

[14] Zachary C Lipton, John Berkowitz, and Charles Elkan. 2015. A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019* (2015).

[15] E Mangalova and E Agafonov. 2014. Wind power forecasting using the k-nearest neighbors algorithm. *International Journal of Forecasting* 30, 2 (2014), 402–406.

[16] T Ozaki. 1977. On the order determination of ARIMA models. *Applied Statistics* (1977), 290–301.

[17] JC Palomares-Salas, JJG De la Rosa, JG Ramiro, J Melgar, A Agüera, and A Moreno. Comparison of models for wind speed forecasting.

[18] Pierre Pinson. 2013. Wind energy: Forecasting challenges for its operational management. *Statist. Sci.* (2013), 564–585.

[19] Richard Piwko and Gary Jordan. 2011. Impacts of Improved Day-Ahead Wind Forecasts on Power Grid Operations. *Contract* 303 (2011), 275–3000.

[20] Ismael Sanchez. 2006. Short-term prediction of wind energy production. *International Journal of Forecasting* 22, 1 (2006), 43–56.

[21] Muhammad Shoaib, Imran Siddiqui, Shafiqur Rehman, Saif Ur Rehman, Shamim Khan, and Aref Lashin. 2016. Comparison of Wind Energy Generation Using the Maximum Entropy Principle and the Weibull Distribution Function. *Energies* 9, 10 (2016), 842.

[22] Lucas Silva. 2014. A feature engineering approach to wind power forecasting: GEFCom 2012. *International Journal of Forecasting* 30, 2 (2014), 395–401.

[23] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*. 3104–3112.

[24] Adrian Trapletti, Friedrich Leisch, and Kurt Hornik. 2000. Stationary and integrated autoregressive neural network processes. *Neural Computation* 12, 10 (2000), 2427–2450.

[25] Qianyao Xu, Dawei He, Ning Zhang, Chongqing Kang, Qing Xia, Jianhua Bai, and Junhui Huang. 2015. A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining. *IEEE Transactions on Sustainable Energy* 6, 4 (2015), 1283–1291.

[26] Erdong Zhao, Jing Zhao, Liwei Liu, Zhongyue Su, and Ning An. 2015. Hybrid wind speed prediction based on a self-adaptive ARIMAX model with an exogenous WRF simulation. *Energies* 9, 1 (2015), 7.