

Mishra, Sambeet; Bordin, Chiara; Taharaguchi, Kota; Palu, Ivo

Article

Comparison of deep learning models for multivariate prediction of time series wind power generation and temperature

Energy Reports

Provided in Cooperation with:

Elsevier

Suggested Citation: Mishra, Sambeet; Bordin, Chiara; Taharaguchi, Kota; Palu, Ivo (2020) : Comparison of deep learning models for multivariate prediction of time series wind power generation and temperature, Energy Reports, ISSN 2352-4847, Elsevier, Amsterdam, Vol. 6, Iss. 3, pp. 273-286,
<https://doi.org/10.1016/j.egy.2019.11.009>

This Version is available at:

<https://hdl.handle.net/10419/243998>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Tmrees, EURACA, 04 to 06 September 2019, Athens, Greece

Comparison of deep learning models for multivariate prediction of time series wind power generation and temperature

Sambeet Mishra^{a,*}, Chiara Bordin^{b,*}, Kota Taharaguchi^a, Ivo Palu^a

^a Tallinn University of Technology, Estonia

^b The Arctic University of Norway, Norway

Received 19 September 2019; received in revised form 29 October 2019; accepted 4 November 2019

Abstract

Wind power experienced a substantial growth over the past decade especially because it has been seen as one of the best ways towards meeting climate change and emissions targets by many countries. Since wind power is not fully dispatchable, the accuracy of wind forecasts is a key element for the electric system operators, as it strongly affects the decision-making processes. The planning horizon can be short term (1–3 months) and long-term (6–12 months) depending on the process.

The objective of this paper is to conduct a performance comparison of five deep learning models each combined with three types of data pre-processing and used for short term and long-term multi-variate predictions. The input data are time series of the wind power capacity factor and the temperature. In addition, this paper sets out to demonstrate and review the state-of-the-art deep learning models for prediction with a secondary objective to present the reader a reference point to better understand which model to choose and what factors are significant. The first contribution of this paper is to apply, assess and compare a selection of the novel and cutting-edge deep learning models for multi-variate prediction. Multi-variate prediction is achieved through a proposed multiple input and multiple output (MIMO) architecture. Compared to traditional prediction models, machine learning techniques have the advantage of generalization. Among various techniques deep learning is particularly getting more attention due to the applicability to various dataset such as numerical and character. This investigation focuses on five models — Deep Feed Forward (DFF), Deep Convolutional Network (DCN), Recurrent Neural Network (RNN), Attention mechanism (Attention) and Long Short-Term Memory Networks (LSTM). The second contribution is to propose a novel approach to transform the time series dataset to signal for input and reconstruct the model predictions through inverse transformation, by means of the so-called discrete wavelet transformation and fast Fourier transformation. The different models are assessed also by comparing their performance with and without the input dataset manipulation through wavelet and FFT transformation. Beyond that, the model performances are outlined in detail, to give the reader an overview of the models to choose from for short-term or long-term prediction. The results demonstrate that the Attention and DCN perform best with Wavelet or FFT signal, whereas some other models perform better with no data preprocessing.

© 2019 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the scientific committee of the Tmrees, EURACA, 2019.

Keywords: Unsupervised machine learning; Multi-variate prediction; Short and long term prediction; Wind power forecasting; Time series to frequency transformation; Performance evaluation and comparison

* Corresponding authors.

E-mail addresses: sambeet.mishra@ttu.ee (S. Mishra), chiara.bordin@uit.no (C. Bordin).

<https://doi.org/10.1016/j.egy.2019.11.009>

2352-4847/© 2019 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the scientific committee of the Tmrees, EURACA, 2019.

1. Introduction

Even if wind energy is intermittent and stochastic in nature, it is increasingly important in the power generation because it is a clean, sustainable and pollution-free source [1]. Since wind power is not fully dispatchable, the accuracy of wind forecasts is a key element for the electric system operators, as it strongly affects the decision-making processes [2]. In fact, wind generation can impact the electricity costs and prices and its potential effects on the overall power system are of interest to power system planners and policy makers globally [3]. The variation in wind power can jeopardize the power system quality and stability as well as affect the market participants who bear the economic losses due to failure of producing the contracted amount of energy [1]. As discussed in [3] day-ahead electricity spot prices are affected by day-ahead wind power forecasts. The relationship between wind power forecasts and spot prices is shown not only to exist, but also to be highly non-linear and time dependent. The spot price is, on average, shown to decrease with increased predicted wind power penetration, while intra-day price variations diminish to some extent. The relationship between intermittent wind power generation and electricity price has been investigated also in [4]. The results show that variable wind power reduces the price level but increases its volatility. The dramatic impact of wind power generation on the overall power system and market outlined in the studies mentioned above, motivates to study the problem of prediction of time-series wind power. There is a need of tools to accurately forecast wind power generation in order to mitigate the undesirable influences of integrating wind energy into the electric power grids. Moreover, several studies showed the impact of weather regime and meteorological data on the wind power ramp forecasts [5,6]. Temperature has been found as an important meteorological factor affecting the wind speed [7,8]. This motivates to study wind power generation forecast by considering not only historical wind data, but also weather data, such as temperature.

Forecasting of wind power generation has therefore received wide attention in literature. Several reviews of different approaches, models and methodologies for wind speed and power prediction can be found in [9–13]. Such works emphasize the diversity of various forecasting methods available and the need for comparisons in order to investigate which approaches perform better, in which conditions and why. No forecasting models can be perfect in every condition, hence more in-depth comparative studies and methodologies assessments should be carried out by the ongoing research. Moreover, novel machine learning techniques should be further investigated, applied and compared to further improve the forecasting approaches within the wind power generation field. The main objective of this paper is to forecast wind power both from historical wind generation data and from meteorological data (namely, temperature), using deep learning models. The first contribution of our paper is to apply, assess and compare a selection of the cutting-edge deep learning models for wind power prediction. Compared to traditional learning methods, deep learning methods have the advantage of data learning and generalization ability. The focus will be given in particular to Long Short-Term Memory Networks (LSTM), Artificial Neural Networks (ANN), Attention based models and Deep Convolutional Networks (DCN). The wind power forecast will be performed by focusing on one particular deep learning modeling structure, namely “Multiple Input, Multiple Output” (MIMO). This means that the deep learning models will be fed with multiple input dataset (wind power and temperature historical values) and they will provide a multiple set of predictions as output (wind power and temperature predictions). The second contribution is to enhance the model’s comparisons by analyzing their performance in forecasting over different time horizons. Four different time horizons for forecast will be assessed, from short term to long term. Performance indexes will be calculated based on established evaluation criteria (Average gap, MSE, RMSE and RMSLE) for each model in order to analyze the performance in forecasting over 1 month, 3 months, 6 months and a year.

The third contribution is to propose a novel approach to manipulate the input dataset for the deep learning models, by using a so-called wavelet transformation and fast Fourier transformation. This will provide a smoother input dataset into the models and therefore will improve the overall computational time compared to the traditional data generation approaches. Of course, smoothing the dataset will also affect the model’s performance in terms of accuracy of predictions. Hence the different models will be assessed also by comparing their performance with and without the input dataset manipulation through wavelet and fast Fourier transformations respectively.

The LSTM approach has been applied for wind power forecast in [14] and compared with neural network and support vector machine models. According to this study, LSTM showed a higher prediction accuracy and greater potential for engineering applications. A wind speed forecasting method based on LSTM can be found also in [15]. The experimental results indicate that the LSTM network has the strong nonlinear processing ability and it is suitable for the non-stationary wind speed forecasting computation. As discussed in [16], the ANN have been widely applied

in literature for forecasting purposes within the wind power generation field. A comparative study considering three ANN is presented in [17]. The results show that even for the same wind dataset, no single neural network model outperforms others universally in terms of all evaluation metrics. Moreover, the selection of the type of neural networks for best performance is also dependent upon the data sources. Five different ANN models are developed and assessed in [7] where results are compared with real data and show a very good accuracy of 70%. Another successful application of ANN for wind power prediction can be found in [18] where high prediction accuracy is shown by the results and validations. Regarding Attention based models there is still very little in literature in terms of wind power forecast applications. However, a good introduction to the Attention based models' specific properties can be found in [19]. Attention mechanisms are also introduced in [20] as deep learning methods able to improve the accuracy on many tasks, particularly within natural language processing and image recognition. The importance of attention mechanisms is also discussed in [21] as a promising method to better address the real-world complexity. In addition, an automatic forecasting of time series data with Multifactor Neural Attention can be found in [22]. The novel methodology achieves a 23.9% improvement of forecasts in comparison to other neural networks proposed for time series forecasting to date. Deep Convolutional Networks have been utilized for wind power predictions. An application can be found in [23] where the results demonstrate that the uncertainties in wind power data can be better learned using the proposed approach and that a competitive performance is obtained. Another successful application is presented in [24] where authors obtained a mean square error reduction by 49.83% compared to state-of-the-art methods. A survey to get an overview of Adversarial Networks applications can be found in [25] where applications in different fields are discussed. However, more focused applications on wind power are discussed in [22,26]. The former addresses wind speed forecasting, while the latter focuses on Wind power scenarios for stochastic optimization problems for power systems in which wind power is a significant component.

The rest of the paper is organized as follows: section *Machine Learning Models for Prediction* will introduce the key properties of the deep learning models addressed in this paper. Section *Input data preparation and transformation* will outline the data pre-processing tasks and their implications. Finally, section *Model Tests and Results* will demonstrate the results obtained after the models' comparison.

2. Machine learning models for prediction

In [27] the authors presented the view that there is a vast literature available on machine learning algorithms, and they presented the advances over existing algorithms. However very little has been done is investigating the relative performance of the models. This paper aims at addressing this issue and provide a comparative overview about models and their performances. The input parameters for the prediction are wind capacity factor and temperature time-series dataset. The predictions can be of univariate or multi-variate type. The former is used in case of single variable predictions and the latter is used for predicting multiple variables that captures the relation between the input parameters. The purpose of this section is to briefly introduce the main properties of the machine learning models addressed in this paper. For prediction a multiple input and multiple output (MIMO) architecture is chosen that takes two parameters, wind capacity factor and temperature, as inputs and performs forecasts as shown in Fig. 1. State-of-the-art deep learning models will be compared based on their performance on predictions, in order to demonstrate and outline their potential and their differences. The following paragraphs will introduce the most important aspects of Deep Feed Forward (DFF), Deep Convolutional Network (DCN), Attention mechanism (Attention) and Long Short-Term Memory Networks (LSTM). Fig. 2 illustrates the schematics of the models considered for this investigation. The models are programmed using TensorFlow [28] package from Google in Python environment [29]. The fundamental principles behind machine learning algorithms can be summarized as in (1) and (2). The (1) reflects the feed forward network where the input value i is multiplied with the weight matrix w [28] while h presents the hidden state. In addition to that a bias b is introduced. Note that in (1), the output is not time dependent, therefore the memory effect (namely the ability to carry forward information from past events) is not taken into account. On the contrary, in (2) the output is time dependent and therefore the memory effect is included [28]. In particular, h_t presents the hidden state at time step t which is a function of the input it altered with a weight matrix W . Moreover, the hidden state of the previous time step h_{t-1} is multiplied with a transition matrix S . The weight refers to the level of importance for present and past hidden states. The errors are returned through back propagation and used to adjust weights until the best performance is achieved. The activation function

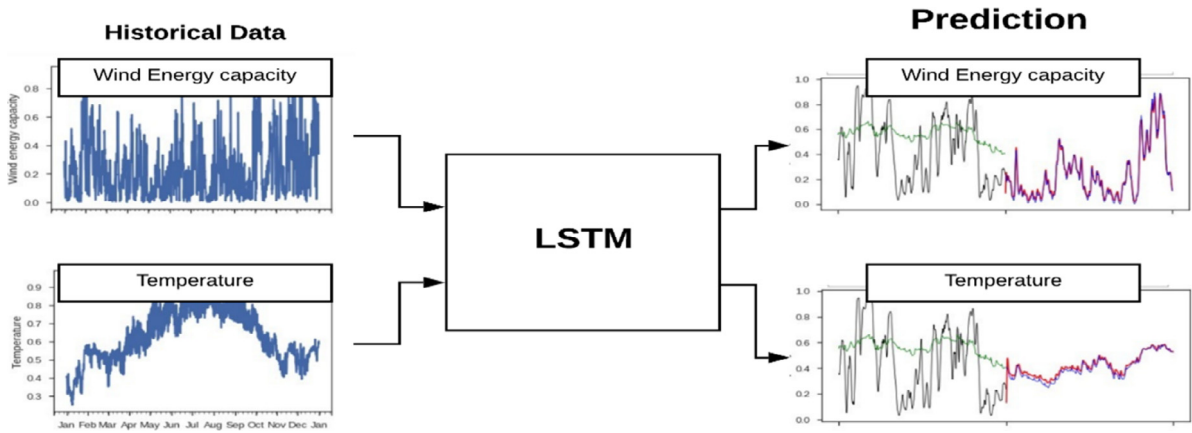


Fig. 1. General structure for predictions.

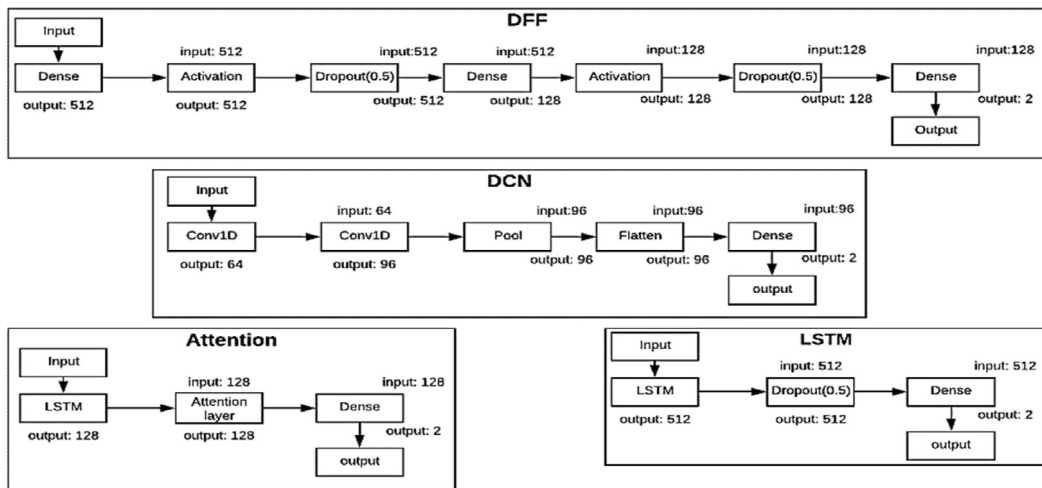


Fig. 2. Model structures and input parameters.

η , groups the sum of different weights and hidden states, through logistic sigmoid function, or tanh or Rectified Linear Unit (ReLU) [30]. For the purpose of this paper, the ReLU is adopted.

$$h = \eta(wi) + b \quad (1)$$

$$h_t = \eta(w \cdot i_t + S \cdot h_{t-1}) \quad (2)$$

Deep convolutional network [31] is a Deep Learning algorithm which can take any input and identify the underlying patterns. While in primitive methods filters are hand-engineered, with enough training, convolutional networks have the ability to learn these characteristics themselves. There are two unidimensional convolutions from the input to the first layer. Then, followed by the form of output, the algorithm builds each layer that shapes the output data. The pooling layer is a layer that performs element compression. DFF [32] consists of neurons, that are ordered into layers which contain input, hidden layer, and output. Each neuron in one layer is connected to every neuron on the next layer. Therefore, information is constantly feed forward from the current state. Firstly, a dense layer can receive the input data and generate an output of 512 dimensions. Next, there is an activation layer which calculates a weighted sum of the input, followed by a dropout layer which is a simple way to prevent neural networks from over-fitting. Lastly, a dense layer adapts the multi outputs which have two dimensions. RNN is a network with feedback loops that allows information to persist. However, the challenge for RNN is long-term dependency. In fact, as the gap between the relevant information and the point where it is needed grows, RNN becomes unable to

learn how to connect the incorrect information. At this point, LSTM is explicitly designed to avoid the long-term dependency problem, with four interacting layers. The LSTM networks are categorized as the kind of RNN which is able to predict long-term dependencies. This idea was introduced in [33]. LSTM has the same structure of RNN, but different modules inside. This can avoid the long-term dependency issues. The significant achievement of LSTM is the ability to make use of short-term memory over a long period. The model consists of three layers (two LSTM layers and one dense layer). These layers can receive the input data and generate an output of 512 dimensions. There is a dropout layer which is a simple way to prevent neural networks from over fitting. Lastly, a dense layer adapts the multi outputs which have two dimensions. The sequence to sequence (seq-to-seq) algorithm is one of the most promising methodologies for prediction in deep learning models. A seq-to-seq model has two components, an encoder and a decoder. The former encodes the input to a context vector, and the latter decodes the context vector with each time step. Attention is a seq-to-seq type mechanism. It was developed in [34] to overcome the issue of fixed context vector, by introducing filtering priority for each time step.

The authors in [35] presented the Adaptive Moment estimation (ADAM) optimizer as an extension to stochastic gradient descent. ADAM is based on adaptive moment estimation. ADAM has two parts — Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp). ADAM computes the individual adaptive learning rates for different parameters, from estimates of first and second moments of the gradient. The Table 1 outlines the input parameters for the models considered for investigation. Standard parameters based on literature were selected for a comparative analysis among the five models. In a future study hyper parameter tuning will be performed wherein the parameters would be optimized for each model for better performance.

Table 1. Parameters selected for the deep learning models for comparison.

Model	Batches	Epoch	Optimizer	Learning rate
Attention	168	200		
DCN	168	200		
DFF	168	200	ADAM	1.00E–03
indRNN	168	200		
LSTM	168	200		

3. Input data preparation and transformation

The wind power production and temperature data for 6 years (2011–2016) are collected from [36] for Paldiski, Estonia. Using the min–max normalization method wind capacity factors (WF) are generated. The Fig. 3, in section (a) and (c), shows the heat map and the three-dimensional plot for temperature and wind power factor input data structure. In particular, for the heat map, the lighter the color is, the higher the temperature or wind values are. It is evident that the temperature is high from June to August due to the influence of the summer season, while the temperature is low from December to February. According to the plot, the temperature rises from April to May in all the years. Moreover, a comparison between January in 2011 and 2016 shows that the temperature was lower in 2011. There has been a little change in temperature year by year since May. It turns out that there is relatively higher production in winter than in summer season. On the other hand, it is evident that the smallest amount of wind power capacity factor is in July for most of the years. From 2011 to 2016, the most significant wind power capacity factor was 0.97, the lowest was 0. The Fig. 3, in section (b) and (d), shows that the capacity factor for December is the highest in any year. Furthermore, in January and March on 2013 and 2014, the capacity factor is lower compared to other years, and the same trend can be seen in May and July. According to the procured data the highest temperature from 2011 to 2016 was 29.47 °C, and the lowest temperature was –30.11 °C. Historical time-series from 2011–2015 have been used as input. A 50–50 division was used for training and testing to avoid over or under fitting problems. Historical data for 2016 was used for validation. For example, January 2016 is used for one-month validation, January to March 2016 was used for 3 months validation. Similarly, January to June 2016 and January to December 2016 were used for 6 months and 12 months validations respectively. The input dataset has been manipulated through wavelet transformation and fast Fourier transformation (FFT). These are methodologies that can be used to transform the data from time to frequency domain. Wavelets are a popular tool for computational analysis. They provide localization in both the temporal domain as well as in the frequency domain [37]. A prominent feature is the ability to perform a multi-resolution analysis [38]. This method is key

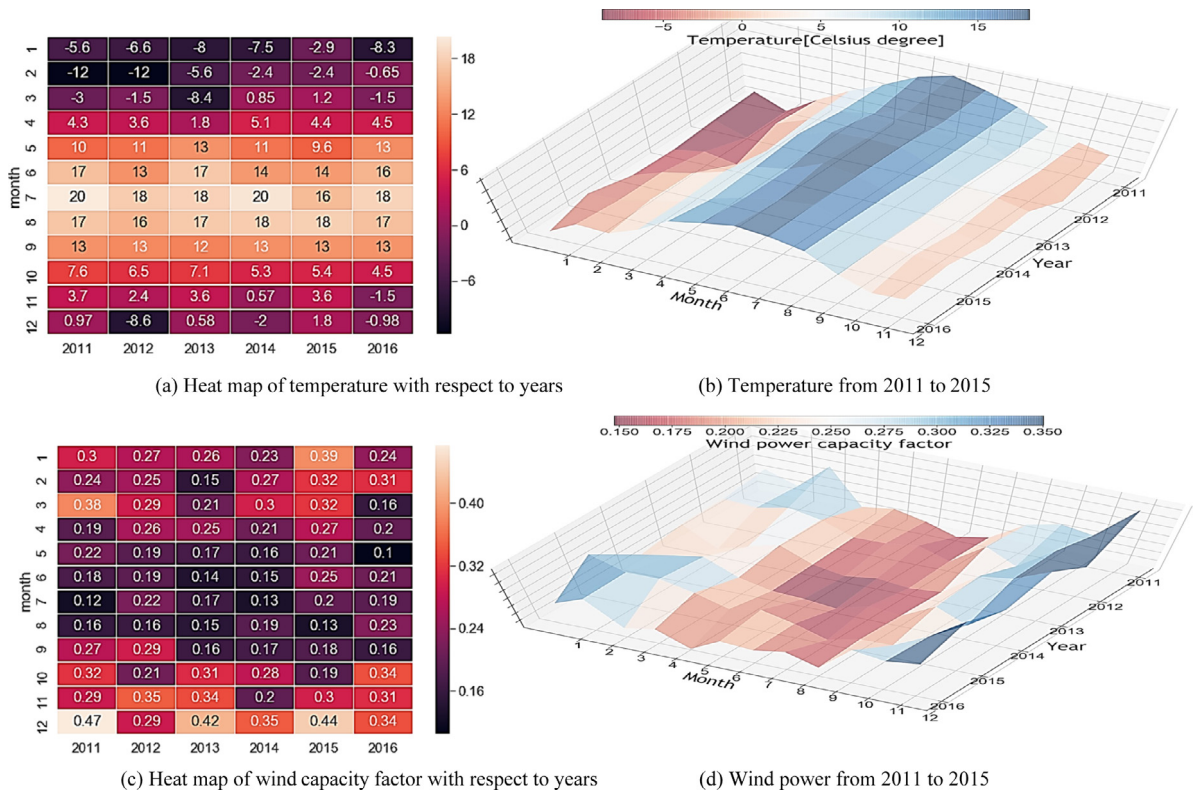


Fig. 3. Input wind power and temperature data set.

to ensure a good performance of wavelets in applications such as data compression and de-noising. The discrete wavelet transform (DWT) operates over scales and positions based on the power of two components. As per [39,40] when decomposing the original signal y_t , two components A_j and D_j are produced by convoluting the signal with a decomposition low pass filter (DLP) and a decomposition high pass filter (DHP) respectively. In (3) y_t presents the actual value at time t , while cA presents the approximation coefficient, and cD stands for detail coefficient.

$$y_t = cA_j + \sum_{j=1}^J cD_j \quad (3)$$

The Fast Fourier transformation (FFT) is applied to transform the time-series data as in [41]. This transformation computes the n -dimensional, n -point discrete Fourier transform with the efficient Fast Fourier Transform algorithm. As shown in Fig. 4, through wavelets and FFT transformations, the dataset can be converted from a time series domain into a frequency domain, which is smoother and therefore easier to process by the deep learning algorithms. Pywavelet package [42,43] for python is used for the transformations. The Scipy package in python provides forward and reverse FFT transformation which was used for this investigation. Fig. 5 gives an overview of the input data processing process, while Fig. 6 shows how wavelet and FFT transformations are placed within the forecasting process. The original time series-based dataset is first converted into a frequency domain signals and fed into the prediction model. The predicted signal is then converted back into a time series domain through inverse transformation. Note that the original time-series data has many variations in contrast with the signal which are smoother by nature.

To better understand the implications of wavelet and FFT transformations, let us look at a prediction example. Fig. 7 shows an example of wind power prediction using indirect recurrent neural networks (IndRNN), when using an original time-series data structure, which means, a data structure that has not been manipulated. The Figs. 8 and 9 shows the prediction results when wavelet and FFT transformation are applied respectively. The performance

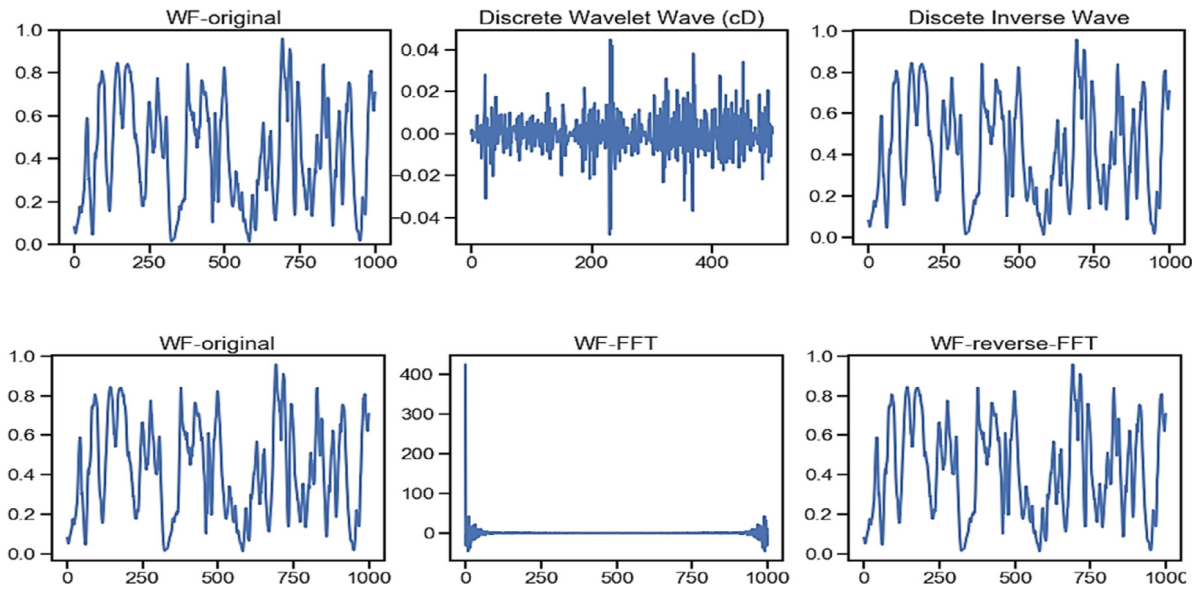


Fig. 4. Sample time-series to signal transformation using discrete wavelet (top) and Fast Fourier Transformation (bottom).

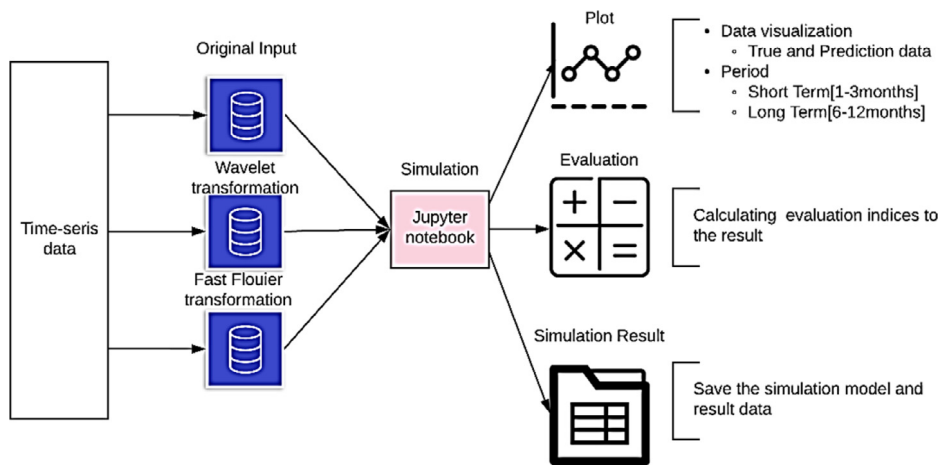


Fig. 5. Schematic representation of the input data processing.

measures including average gap, mean average percentage error, mean square error, mean square root error and logarithmic mean square root error are presented for 1, 3, 6- and 12-months prediction period respectively. The figures show that there is a loss in reconstruction of signals from frequency domain to time-series. In fact, the two curves (namely the prediction curve in red and the actual curve in blue) visually look similar, but there is a slight gap when wavelet and FFT transformation are applied. A detailed comparison of model performances is presented in the following section.

The Table 2 presents the machine specifications used for the purpose of this investigation. To avoid over or under fitting the input dataset is divided in 50–50 ratio for training and validation purpose. The execution times are also recorded to demonstrate the reader what to expect while performing similar tests. The models are saved in Hierarchical Data Format (HDF) for re-usage.

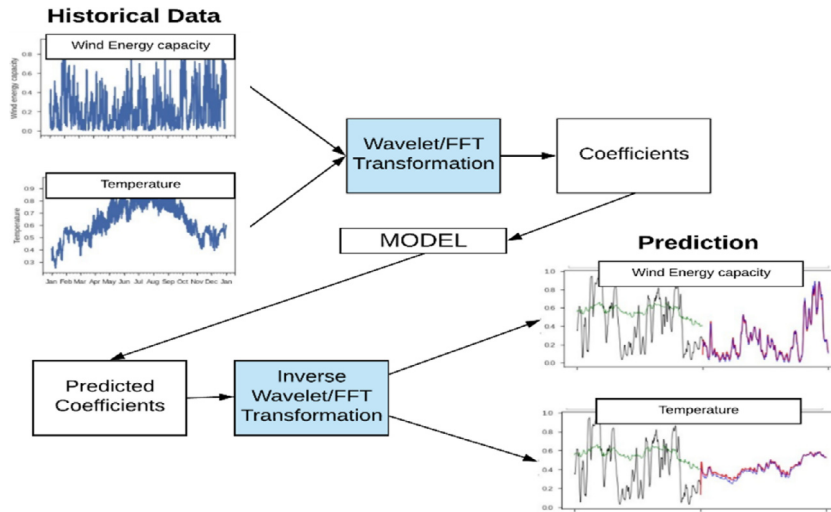


Fig. 6. Insertion of data transformation within the pre-processing and post-processing operations.

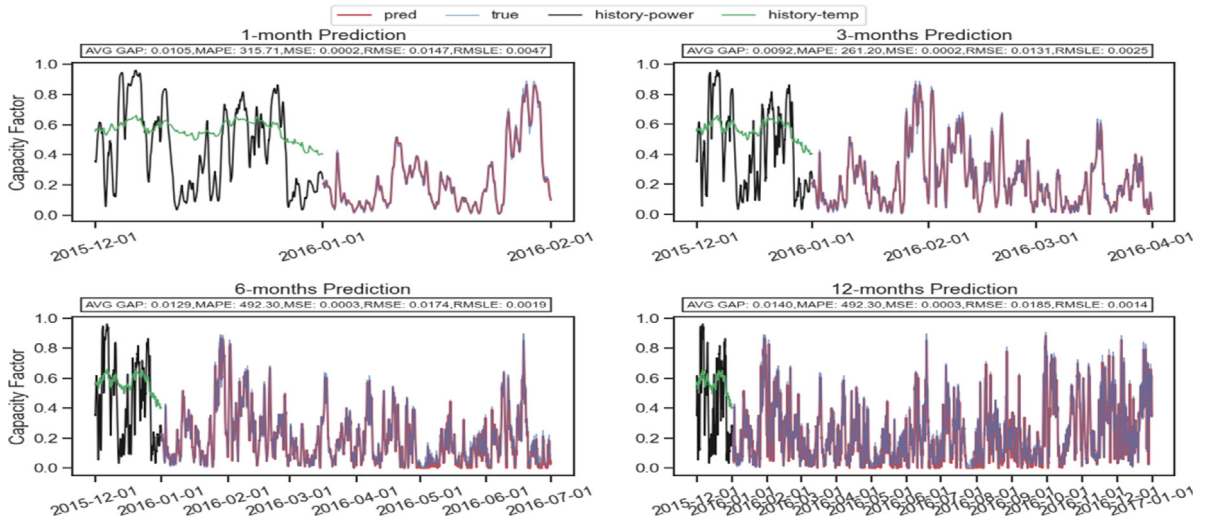


Fig. 7. Prediction using indRNN using original input data.

Table 2. Machine specifications where the simulation was run.

Windows edition	Windows 10 Pro
Processor	Intel i7 CPU@3.4 GHz
Installed memory (RAM)	16 GB
System type	64-bit operating system

4. Model tests and results

The selected deep learning models have been tested for wind and temperature prediction, with and without wavelet and FFT transformation and over different time horizons, from short term to long term. The objective is to get insight in the overall performance of the different models and investigate their suitability in different conditions. It is possible to note that a common machine specification with average computational capabilities has been used. Therefore, the proposed models can perform well with conventional machines and there is no need to use special computational power or expensive IT infrastructures. However, finding optimal parameters for the model to improve

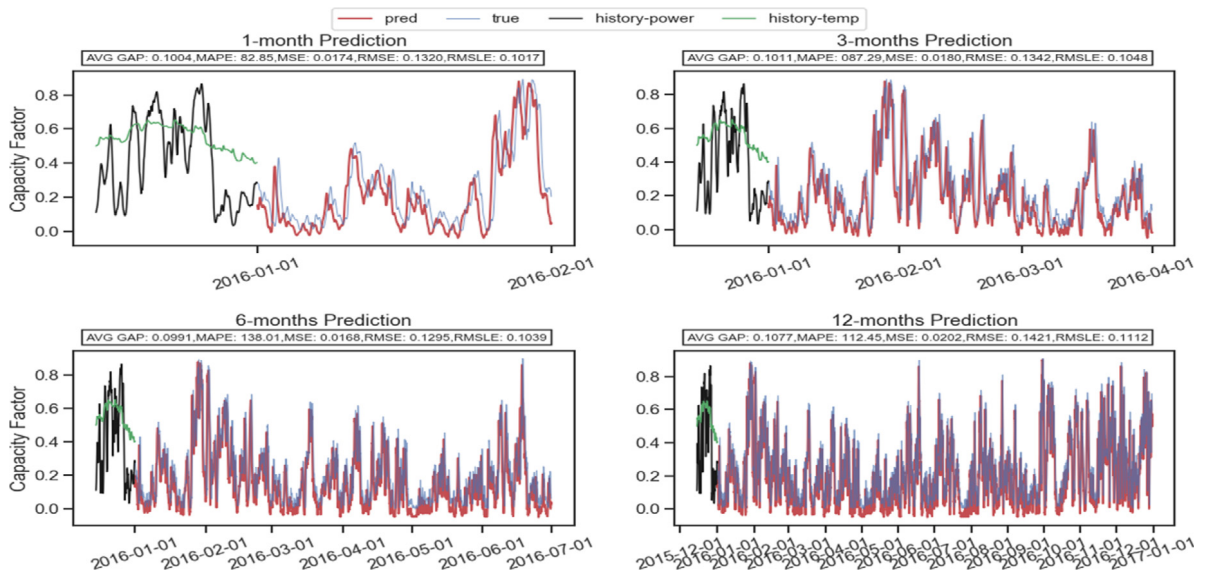


Fig. 8. Prediction using IndRNN with pre-processing using wavelet transformation.

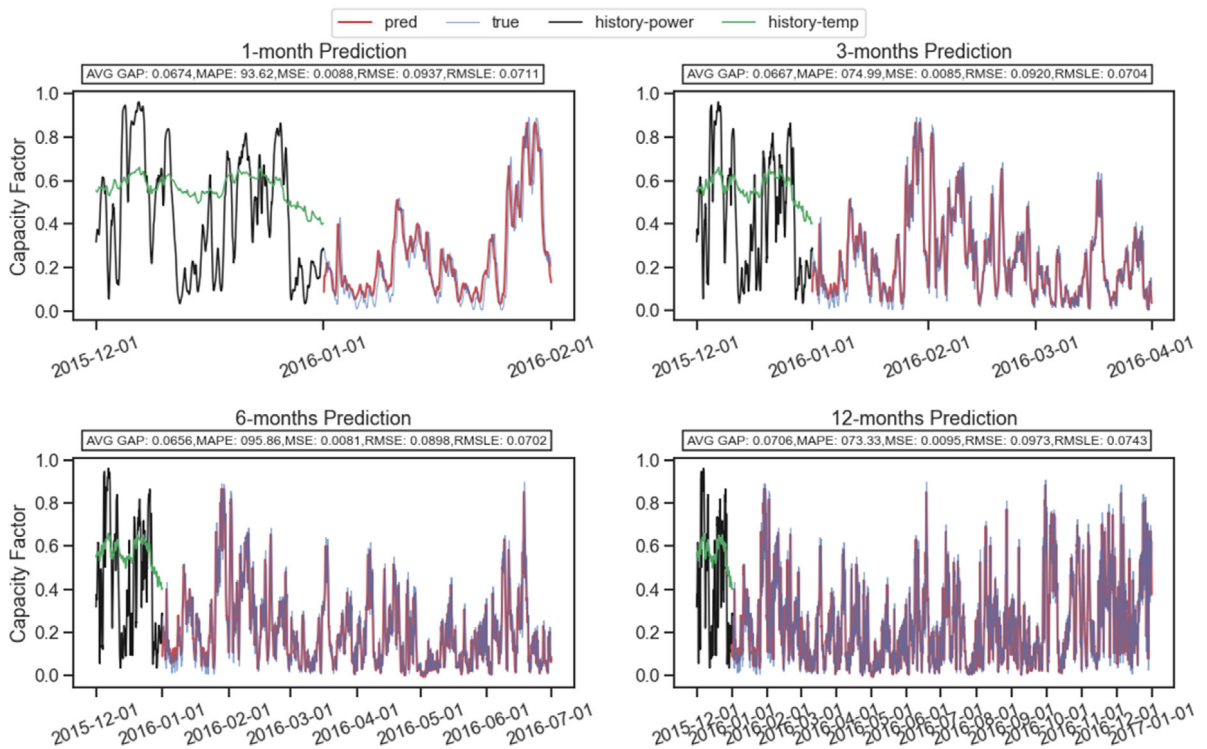


Fig. 9. Prediction using IndRNN with pre-processing using FFT.

the performance, might require a more powerful machine set-up. In addition, if a very high volume of data has to be trained then more powerful and expensive infrastructures may be needed. The following paragraphs will present and describe the prediction results of each model for wind power capacity factor and temperature. Each result included short-term (1 and 3 months) and long-term (6 and 12 months) forecasts. For evaluating the model performance, the

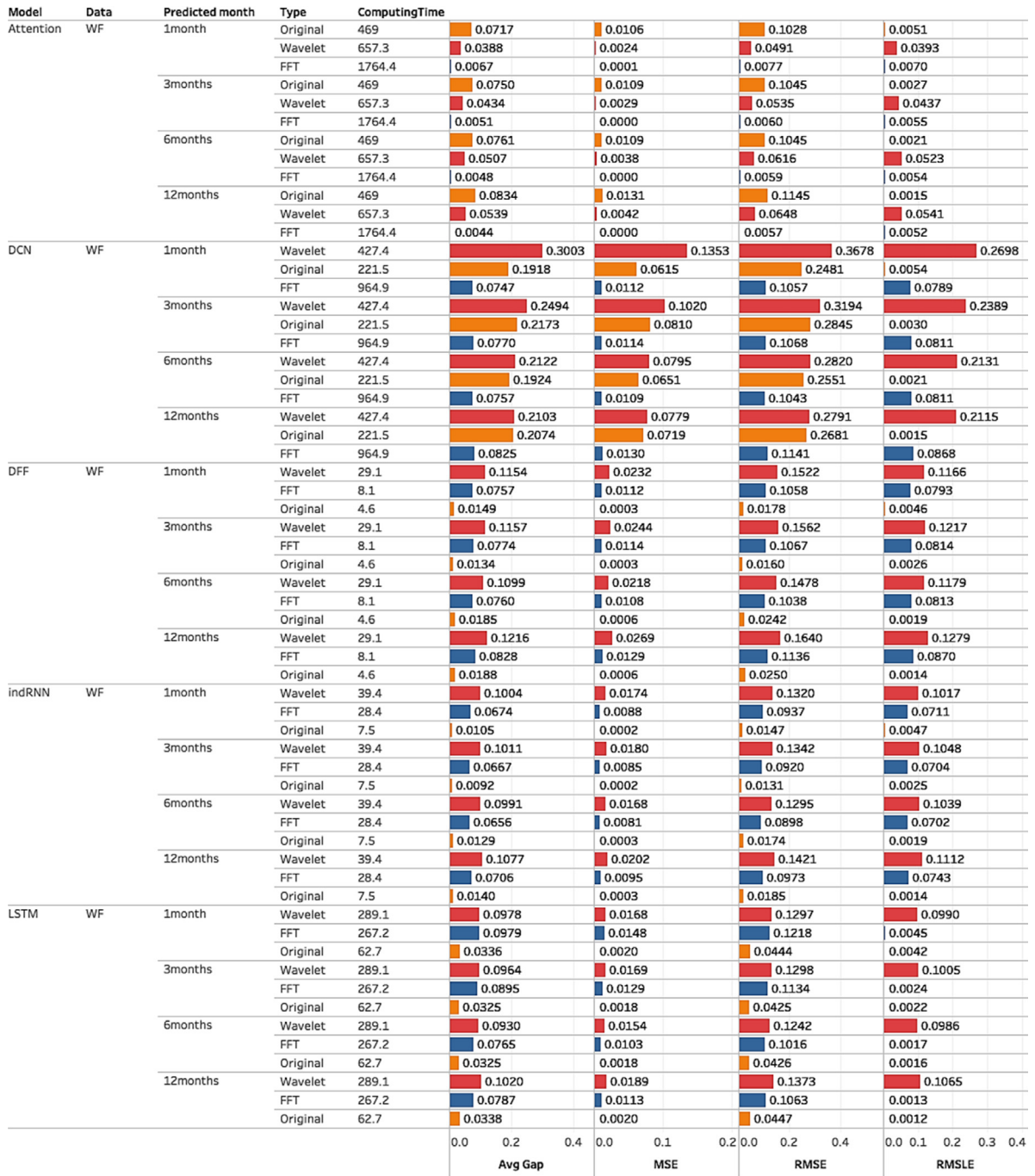


Fig. 10. Performance table based on error metrics for the models with and without pre-processing using Wind capacity Factor (WF) as input.

following evaluation criteria are considered — average gap (AG), Mean Square Error (MSE), Root Mean Square Error (RMSE), Root Mean Squared Log Error (RMSLE). Eqs. (4) to (7) present the definitions of each evaluation indices for AG, MSE, RMSE and RMSLE respectively. MSE incorporates both the variance and the bias of the predictor. RMSE is the square root of MSE. In case of unbiased estimator, RMSE is the square root of variance, which is Standard Deviation. RMSLE takes the log of the predictions and actual values. RMSLE is particularly used

Model	Data	Predicted month	Type	ComputingTime	Avg Gap	MSE	RMSE	RMSLE
Attention	TEMP	1month	Original	469	0.0221	0.00070	0.0267	0.00170
			Wavelet	657.3	0.0107	0.00020	0.0138	0.00940
			FFT	1764.4	0.0080	0.00010	0.0088	0.00650
		3months	Original	469	0.0274	0.00120	0.0352	0.00150
			Wavelet	657.3	0.0152	0.00040	0.0191	0.01260
			FFT	1764.4	0.0050	0.00000	0.0065	0.00460
		6months	Original	469	0.0438	0.00320	0.0570	0.00190
			Wavelet	657.3	0.0233	0.00090	0.0302	0.01810
			FFT	1764.4	0.0166	0.00060	0.0235	0.01340
		12months	Original	469	0.0404	0.00290	0.0535	0.00150
			Wavelet	657.3	0.0254	0.00100	0.0316	0.01850
			FFT	1764.4	0.0198	0.00070	0.0269	0.01520
DCN	TEMP	1month	Wavelet	427.4	0.1056	0.01440	0.1200	0.0837
			Original	221.5	0.0949	0.01320	0.1150	0.00160
			FFT	964.9	0.0183	0.00060	0.0241	0.01720
		3months	Original	221.5	0.1056	0.01950	0.1398	0.00130
			Wavelet	427.4	0.0720	0.00840	0.0919	0.06330
			FFT	964.9	0.0270	0.00130	0.0367	0.02450
		6months	Original	221.5	0.0825	0.01250	0.1119	0.00190
			Wavelet	427.4	0.0769	0.00950	0.0974	0.06160
			FFT	964.9	0.0448	0.00340	0.0583	0.03500
		12months	Wavelet	427.4	0.0725	0.00830	0.0913	0.05660
			Original	221.5	0.0650	0.00810	0.0899	0.00170
			FFT	964.9	0.0415	0.00290	0.0538	0.03180
DFF	TEMP	1month	Wavelet	29.1	0.0318	0.00150	0.0393	0.02810
			FFT	8.1	0.0225	0.00090	0.0296	0.02110
			Original	4.6	0.0151	0.00040	0.0212	0.00180
		3months	Wavelet	29.1	0.0455	0.00320	0.0563	0.03800
			FFT	8.1	0.0282	0.00140	0.0378	0.02540
			Original	4.6	0.0123	0.00030	0.0167	0.00140
		6months	Wavelet	29.1	0.0566	0.00590	0.0830	0.05070
			FFT	8.1	0.0454	0.00340	0.0586	0.03530
			Original	4.6	0.0183	0.00050	0.0227	0.00190
		12months	Wavelet	29.1	0.0641	0.00630	0.0792	0.04780
			FFT	8.1	0.0422	0.00290	0.0543	0.03220
			Original	4.6	0.0210	0.00070	0.0256	0.00150
indRNN	TEMP	1month	FFT	28.4	0.0689	0.00650	0.0803	0.05800
			Wavelet	39.4	0.0257	0.00100	0.0318	0.02260
			Original	7.5	0.0228	0.00060	0.0251	0.00160
		3months	FFT	28.4	0.0436	0.00330	0.0573	0.04020
			Wavelet	39.4	0.0383	0.00230	0.0475	0.03200
			Original	7.5	0.0172	0.00040	0.0206	0.00150
		6months	Wavelet	39.4	0.0577	0.00520	0.0719	0.04380
			FFT	28.4	0.0463	0.00330	0.0578	0.03730
			Original	7.5	0.0208	0.00070	0.0256	0.00190
		12months	Wavelet	39.4	0.0551	0.00460	0.0680	0.04090
			FFT	28.4	0.0405	0.00260	0.0508	0.03190
			Original	7.5	0.0223	0.00080	0.0274	0.00150
LSTM	TEMP	1month	FFT	267.2	0.1865	0.03610	0.1899	0.00560
			Original	62.7	0.0269	0.00110	0.0332	0.00160
			Wavelet	289.1	0.0269	0.00100	0.0316	0.02250
		3months	FFT	267.2	0.1493	0.02380	0.1543	0.00150
			Wavelet	289.1	0.0352	0.00180	0.0429	0.02890
			Original	62.7	0.0202	0.00070	0.0259	0.00150
		6months	FFT	267.2	0.0991	0.01370	0.1169	0.00190
			Wavelet	289.1	0.0520	0.00420	0.0646	0.03920
			Original	62.7	0.0247	0.00100	0.0316	0.00190
		12months	FFT	267.2	0.0872	0.01120	0.1059	0.00140
			Wavelet	289.1	0.0497	0.00380	0.0615	0.03680
			Original	62.7	0.0235	0.00090	0.0307	0.00150

Fig. 11. Performance table based on error metrics for the models with and without pre-processing using temperature (TEMP) as input.

to avoid penalizing huge differences in the predicted and the actual values when both predicted and true values are huge numbers. If both predicted and actual values are small, then RMSE and RMSLE is same. If either the predicted or the actual values are big then $RMSE > RMSLE$. If both the predicted and the actual values are big the $RMSE > RMSLE$, thus RMSLE becomes negligible [44]. In Eqs. (4) to (7), n is the number of time points; $(y_t - \hat{y}_t)^2$ refers to the squared differences, y_t is the actual value at time t ; \hat{y}_t is the predicted value at time t .

$$AG = \frac{1}{n} \sum_{i=1}^n |(y_t - \hat{y}_t)| \tag{4}$$

$$MSE = 1/n \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

$$RMSLE = \sqrt{\frac{1}{n} \sum_{j=1}^n (\log(y_j + 1) - \log(\hat{y}_j + 1))^2} \quad (7)$$

Figs. 10 and 11 presents a graphical overview of the performance of five models covering short- and long-term predictions for wind capacity factor and temperature respectively. The computing time for each model is captured and presented on seconds resolution. Note that, Attention has performed overall better in both short- and long-term prediction followed by Deep Feed Forward (DFF). This implies that recurrent neural networks perform better in time-series prediction tasks due to the feedback adjustment over the deep neural network. Furthermore, Attention performs better when wavelet transform is applied to the input data set. While Deep Convolutional Network (DCN) performs better when FFT transformation is applied to the input dataset. As for the other deep learning algorithms, they perform better when the input data are preserved in the time-series. Note that the transformation from time series to frequency domain, improves the results in almost all the cases. However, looking at the RMSLE index the performances vary. This is due to the fact that RMSLE considers logarithm scale that is more minute than the decimal system. Hence RMSLE captures the minute changes thereby capturing the over and under fitting. Looking at Fig. 10, where the performance results for wind power forecast are shown, the following observations can be made. Regarding the Attention mechanism, for both short-term and long-term prediction, the FFT data manipulation gives better forecast, followed by the wavelet transformation. Regarding DCN, for short-term predictions, the FFT data manipulation gives better forecast, followed by the original dataset. Hence in this case wavelet transformation gives less accurate results. However, as the time horizon increases towards long-term predictions, it can be observed that the wavelet data manipulation gives as good forecasts as the original dataset. Hence, given that wavelet is improving the computational time, it is worthy it to be used, because its performance is almost the same as the performance that we would obtain by using the dataset as is. Regarding DFF, indRNN and LSTM, for both short-term and long-term prediction, the original dataset is giving more accurate results compared to a dataset manipulated with wavelet or FFT. Looking at Fig. 11, where the performance results for temperature forecast are shown, the following observations can be made. In general, the overall performance when predicting temperature look better than the overall performance observed for wind power forecasts. This might be due to the intrinsic intermittent nature of wind whose frequent variations create more forecasting challenges. Regarding the Attention mechanism, for both short-term and long-term prediction, still the FFT data manipulation gives better forecast, followed by the wavelet transformation. Regarding DCN, for short-term predictions, the FFT data manipulation gives better forecast, followed by the original dataset. Hence in this case as well, the wavelet transformation gives less accurate results. However, as the time horizon increases towards long-term predictions, it can be observed that the wavelet data manipulation gives as good forecasts as the original dataset. This is a result similar to the one previously obtained for the wind power forecast. Regarding DFF, indRNN and LSTM, for both short-term and long-term prediction, the original dataset is giving more accurate results compared to a dataset manipulated with wavelet or FFT. For DFF, the worse performance is obtained when using wavelet transformation. Moreover, the overall performance get worse as the time horizon increases. For IndRNN the worse performance is obtained when using FFT transformation for short term predictions while long term predictions get worse when using wavelet transformation. Compared to the wind predictions, when it comes to temperature predictions it can be observed that, for LSTM, the FFT transformation is giving a much worse performance. While for short term predictions the wavelet transformation performs as good as the original dataset.

5. Conclusion

In this paper, a systematic combination of deep learning methods with data preprocessing for short/long-term multi-variate predictions is investigated. The data is pre-processed from time-series to signal transformation using

discrete wavelet and fast Fourier transformation. Multiple prediction horizons ranging from short to long term are considered to better understand the model performances. Moreover, the effect of input data manipulation through wavelet and FFT transformation have been investigated. The results show that Attention mechanisms and DCN perform better when wavelet or FFT transformation are applied to the dataset. Therefore, the dataset transformation from a time series domain to a frequency domain is worth it to be used as it affects positively the computational time and overall performance of the models. For the other models, in several cases the wavelet and FFT transformation performance was as good as the original dataset performance. In general, it was observed that such data manipulation can smooth the dataset and improve the computational time, without worsening the overall forecast accuracy. It is therefore a technique that is worthy it to be further investigated and used within the mentioned machine learning approaches.

The results demonstrated that the overall performance when predicting temperature look better than the overall performance observed for wind power forecasts. This might be due to the intrinsic intermittent nature of wind whose frequent variations create more forecasting challenges. In general, the machine learning models analyzed so far showed very promising results within the field of wind and temperature forecasts, especially when wavelet and FFT transformation were applied. Future research directions should therefore focus more on further investigations and assessment of such methodologies within other energy fields that may be of interests for the scientific community. The machine learning models analyzed, showed very promising results within the field of wind and temperature forecasts, especially when wavelet and FFT transformation were applied. Beyond that there is a scope to investigate and improve the losses in signal reconstruction. In turn, it would improve the prediction accuracy. The future direction of this work would delve into hyper parameter tuning to understand better how change in parameters such as dimensions of different layers affect the accuracy.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] Renani ET, Elias MFM, Rahim NA. Using data-driven approach for wind power prediction: A comparative study. *Energy Convers Manag* 2016;118:193–203.
- [2] Mauch B, Apt J, Carvalho PMS, Small MJ. An effective method for modeling wind power forecast uncertainty. *Energy Syst* 2013;4:393–417.
- [3] Jónsson T, Pinson P, Madsen H. On the market impact of wind energy forecasts. *Energy Econ* 2010;32:313–20.
- [4] Ketterer JC. The impact of wind power generation on the electricity price in Germany. *Energy Econ* 2014;44:270–80.
- [5] Couto A, Costa P, Rodrigues L, Lopes VV, Estanqueiro A. Impact of weather regimes on the wind power ramp forecast in Portugal. *IEEE Trans Sustain Energy* 2015;6:934–42.
- [6] Okada M, Ichizawa T, Nakamura Y, Yamaguchi K, Kodama R, Kato H, et al. Development of a wind power ramp forecast system by a statistical and meteorological approach. In: *Gd. Renew. Energy Proc. Japan Counc. Renew. Energy*. 2018, p. 130.
- [7] Kaur T, Kumar S, Segal R. Application of artificial neural network for short term wind speed forecasting. In: *2016 Bienn. Int. Conf. Power Energy Syst. Towar. Sustain. Energy*. 2016, p. 1–5.
- [8] Li Y, Dai C, Wang T, Zhou Z, Zhou S, Cai L, et al. Separate wind power and ramp predictions based on meteorological variables and clustering method. In: *2016 IEEE 6th Int. Conf. Power Syst.*. 2016, p. 1–6.
- [9] Soman SS, Zareipour H, Malik O, Mandal P. A review of wind power and wind speed forecasting methods with different time horizons. In: *North Am. Power Symp.* 2010. 2010, p. 1–8.
- [10] Wang X, Guo P, Huang X. A review of wind power forecasting models. *Energy Procedia* 2011;12:770–8.
- [11] Foley AM, Leahy PG, Marvuglia A, McKeogh EJ. Current methods and advances in forecasting of wind power generation. *Renew Energy* 2012;37:1–8.
- [12] Jung J, Broadwater RP. Current status and future advances for wind speed and power forecasting. *Renew Sustain Energy Rev* 2014;31:762–77.
- [13] Tascikaraoglu A, Uzunoglu M. A review of combined approaches for prediction of short-term wind speed and power. *Renew Sustain Energy Rev* 2014;34:243–54.
- [14] Xiaoyun Q, Xiaoning K, Chao Z, Shuai J, Xiuda M. Short-term prediction of wind power based on deep long short-term memory. In: *2016 IEEE PES Asia-Pacific Power Energy Eng. Conf.*. 2016, p. 1148–52.
- [15] Liu H, Mi X, Li Y. Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network. *Energy Convers Manag* 2018;156:498–514.
- [16] Ata R, et al. Artificial neural networks applications in wind energy systems: a review. *Renew Sustain Energy Rev* 2015;49.
- [17] Li G, Shi J. On comparing three artificial neural networks for wind speed forecasting. *Appl Energy* 2010;87:2313–20.

- [18] Ramasamy P, Chandel SS, Yadav AK. Wind speed prediction in the mountainous region of India using an artificial neural network model. *Renew Energy* 2015;80:338–47.
- [19] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: *Adv. Neural Inf. Process. Syst.*. 2017, p. 5998–6008.
- [20] Biran O, Cotton C. Explanation and justification in machine learning: A survey. *IJCAI-17 Work. Explain. AI*, Vol. 8, 2017.
- [21] Niv Y, Daniel R, Geana A, Gershman SJ, Leong YC, Radulescu A, et al. Reinforcement learning in multidimensional environments relies on attention mechanisms. *J Neurosci* 2015;35:8145–57.
- [22] Khodayar M, Wang J, Manthouri M. Interval deep generative neural network for wind speed forecasting. *IEEE Trans Smart Grid* 2018.
- [23] Wang H, Li G, Wang G, Peng J, Jiang H, Liu Y. Deep learning based ensemble approach for probabilistic wind power forecasting. *Appl Energy* 2017;188:56–70.
- [24] Yu R, Liu Z, Li X, Lu W, Ma D, Yu M, et al. Scene learning: Deep convolutional networks for wind power prediction by embedding turbines into grid space. *Appl Energy* 2019;238:249–57.
- [25] Zhu K, Liu X, Yang H. A survey of generative adversarial networks. In: *2018 Chinese Autom. Congr.*. 2018, p. 2768–73.
- [26] Jiang C, Mao Y, Chai Y, Yu M, Tao S. Scenario generation for wind power using improved generative adversarial networks. *IEEE Access* 2018;6:62193–203.
- [27] Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and machine learning forecasting methods: Concerns and ways forward. *PLoS One* 2018;13. e0194889.
- [28] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. Tensorflow: A system for large-scale machine learning. In: *12th USENIX Sys. Oper. Syst. Des. Implement.* (2016), p. 265–83.
- [29] Van Rossum G, Drake FL. *The Python Language Reference Manual*. Network Theory Ltd.; 2011.
- [30] Xu B, Wang N, Chen T, Li M. Empirical evaluation of rectified activations in convolutional network. 2015, ArXiv Prepr [ArXiv:1505.00853](https://arxiv.org/abs/1505.00853).
- [31] Dong C, Loy CC, He K, Tang X. Learning a deep convolutional network for image super-resolution. In: *Eur. Conf. Comput. Vis.*. 2014, p. 184–99.
- [32] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: *Proc. Thirteen. Int. Conf. Artif. Intell. Stat.*. 2010, p. 249–56.
- [33] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9:1735–80.
- [34] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. 2014, ArXiv Prepr [ArXiv:1409.0473](https://arxiv.org/abs/1409.0473).
- [35] Kingma DP, Ba J. Adam: A method for stochastic optimization. 2014, ArXiv Prepr [ArXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [36] Staffell I, Pfenninger S. Using bias-corrected reanalysis to simulate current and future wind power output. *Energy* 2016;114:1224–39.
- [37] Daubechies I, Sweldens W. Factoring wavelet transforms into lifting steps. *J Fourier Anal Appl* 1998;4:247–69.
- [38] Mallat SG. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans Pattern Anal Mach Intell* 1989;11:674–93. <http://dx.doi.org/10.1109/34.192463>.
- [39] Nikolaou NG, Antoniadis IA. Application of Wavelet Packets in bearing fault diagnosis. In: *Proc. 5th WSES Int. Conf. Circuits, Syst. Commun. Comput. (CSCC 2001)*, Rethymno, Greece, 2001, p. 12–9.
- [40] Torrésani B. An overview of wavelet analysis and time-frequency analysis (a minicourse). In: *Int. Work. Self-Similar Syst.*. 1998, p. 9–34.
- [41] Mathieu M, Henaff M, LeCun Y. Fast training of convolutional networks through ffts. 2013, ArXiv Prepr [ArXiv:1312.5851](https://arxiv.org/abs/1312.5851).
- [42] Lee G, Wasilewski F, Gommers R, Wohlfahrt K, O’Leary A, Nahrstaedt H. *PyWavelets: Wavelet transforms in python*. 2006.
- [43] Wasilewski F, et al. *PyWavelets-wavelet transforms in python*. 2010.
- [44] Brink H, Richards JW, Fetherolf M, Cronin B. *Real-world machine learning*. Manning; 2017.