

Decomposition-Based Hybrid Models for Very Short-Term Wind Power Forecasting [†]

Juan Manuel González Sopena ^{1,*} , Vikram Pakrashi ^{2,3,4}  and Bidisha Ghosh ^{1,5} 

¹ QUANT Group, Department of Civil, Structural and Environmental Engineering, Trinity College Dublin, D02 PN40 Dublin, Ireland; bghosh@tcd.ie

² UCD Centre for Mechanics, Dynamical Systems and Risk Laboratory, School of Mechanical & Materials Engineering, University College Dublin, D04 V1W8 Dublin, Ireland; vikram.pakrashi@ucd.ie

³ SFI MaREI Centre, University College Dublin, D04 Dublin, Ireland

⁴ The Energy Institute, University College Dublin, D04 V1W8 Dublin, Ireland

⁵ CONNECT: SFI Research Centre for Future Networks & Communications, Trinity College Dublin, D02 Dublin, Ireland

* Correspondence: gonzlezj@tcd.ie

† Presented at the 7th International Conference on Time Series and Forecasting, Gran Canaria, Spain, 19–21 July 2021.

Abstract: Wind power forecasting is a tool used in the energy industry for a wide range of applications, such as energy trading and the operation of the grid. A set of models known as decomposition-based hybrid models have stood out in recent times due to promising results in terms of performance. As many publications on this matter are found in the literature, a comparison of these models is difficult, because they are tested under different conditions in terms of data, prediction horizon, and time resolution. In this paper, we provide a comparison unifying these parameters using the main decomposition algorithms and a set of artificial neural network-based models for very short-term wind power forecasting (up to 30 min ahead). For this purpose, a case study using data from an Irish wind farm is performed to analyze the models in terms of accuracy and robustness for a variety of wind power generation scenarios.

Keywords: very short-term wind power forecasting; decomposition-based hybrid models; artificial neural networks; data-driven forecasting models



Citation: Sopena, J.M.G.; Pakrashi, V.; Ghosh, B. Decomposition-Based Hybrid Models for Very Short-Term Wind Power Forecasting. *Eng. Proc.* **2021**, *5*, 39. <https://doi.org/10.3390/engproc2021005039>

Published: 7 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Wind power forecasting (WPF) is a tool of importance for practitioners in the wind energy industry, and it accomplishes different tasks depending on the time horizon, from reserve requirement decisions [1] to energy trading [2].

Several standards are found in the literature to classify WPF models with respect to the forecast horizon. One of the most well-known conventions is presented in [3], in which four time horizons are defined: very short-term (up to 30 min ahead), short-term (from 30 min to 6 h ahead), medium-term (up to 1 day ahead), and long-term (more than 1 day ahead) horizons. For medium- and long-term forecasts, physical models are preferred, whereas statistical models are used for very short- and short-term horizons, as they are easier to model and less computationally expensive than physical-based approaches. Among the statistical models, a family of models known as decomposition-based hybrid models has gained the attention of wind forecasting practitioners, with more than 100 papers on this topic having been published [4]. These models have a preprocessing step in which the complexity of wind power time series is avoided by decomposing the signal into a set of more stationary components (usually known as modes). However, as the literature on this type of models is already very extensive, it is difficult to determine which of these models are more suitable for very short-term and short-term forecasts, as they are tested under datasets of different nature, length, and resolution. In addition, the resulting components

are usually fit using a broad variety of artificial neural networks (ANNs), whose capacity to identify and model the features of wind power time series differ depending on the intrinsic characteristics of the type of ANN. Taking all these aspects into consideration, the aim of this article is to provide a case study where (1) the state-of-the-art decomposition techniques are considered to decompose wind power time series, (2) a set of ANN models are used to train the resulting modes, and (3) a time-scale classification of the models for very short-term wind power forecasts using common criteria is presented.

The paper is organized as follows: Section 2 introduces the main elements of decomposition-based hybrid models; Section 3 describes the data used in this study; Section 4 presents the results; and Section 5 provides the concluding remarks of this paper.

2. Methodology

In this section, the main decomposition algorithms and ANN-based forecasting models are described, as well as the metrics used to analyze the performance of the models.

2.1. Decomposition-Based Hybrid Models

Decomposition-based hybrid models decompose the original time series into a set of more stationary modes that are easier to handle. In terms of forecasting, ANNs allow us to exploit diverse features of the data, such as recurrent neural networks (RNN) or convolutional neural networks (CNN). The main structure for this family of models is shown in Figure 1: (1) the wind power time series is decomposed into a set of modes; (2) a forecasting model is built independently for every mode; and (3) the wind power forecast is estimated by adding the values of all modes.

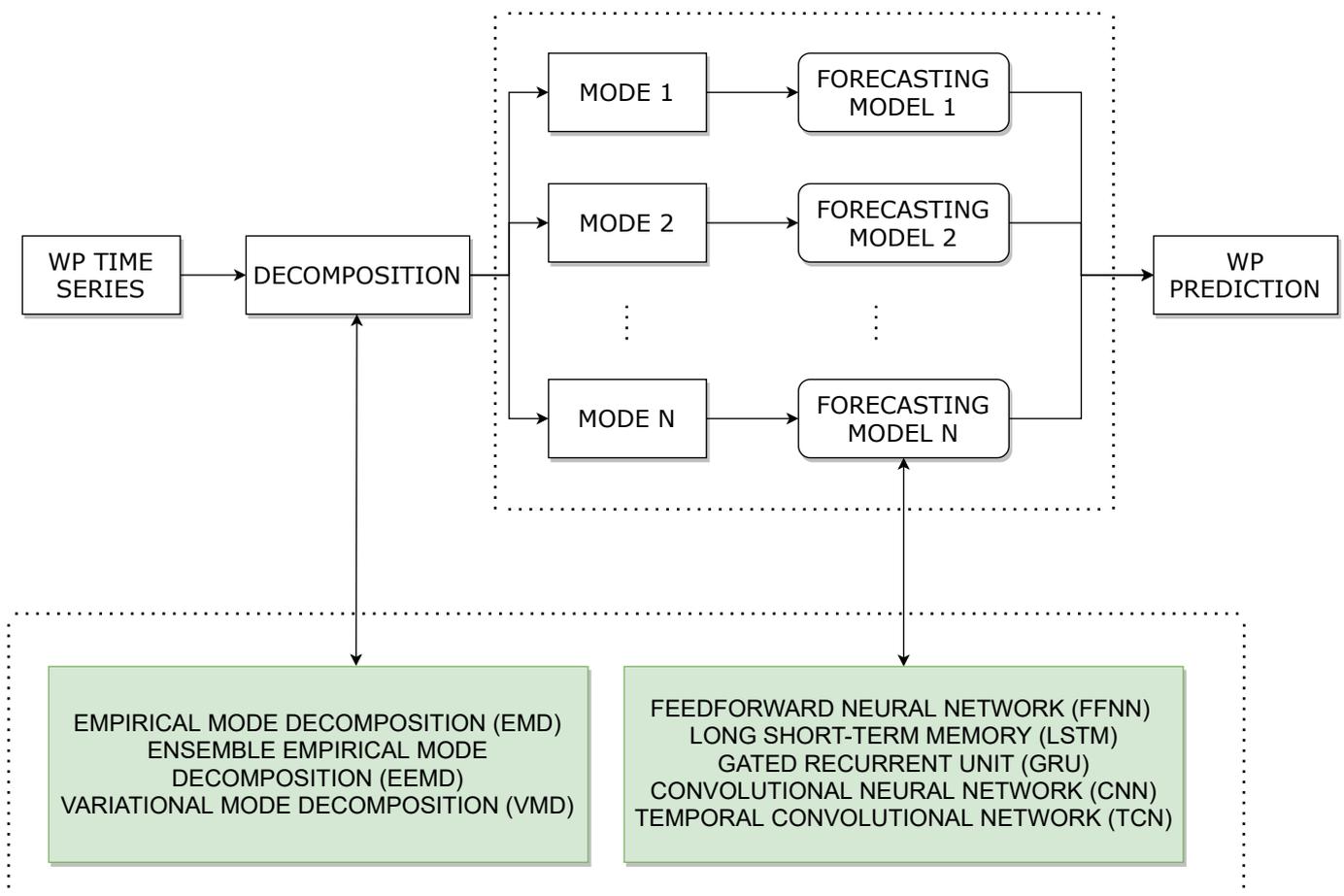


Figure 1. Flowchart for decomposition-based hybrid models.

Two of the most common decomposition techniques are empirical mode decomposition (EMD) [5] and variational mode decomposition (VMD) [6]. The wind power time series are decomposed into modes as

$$y(t) = \sum_{k=1}^K y_k(t) \tag{1}$$

where $y_k(t)$ is the k -mode extracted from the data. The modes, also known in the literature as intrinsic mode functions (IMFs), can be expressed as amplitude-modulated–frequency-modulated (AM–FM) signals [6,7]:

$$y_k(t) = A_k(t) \cos \phi_k(t), \quad A_k(t), \phi'_k > 0 \quad \forall t \tag{2}$$

where $\phi_k(t)$ is a non-decreasing function. The main assumption is that the variation of A_k and ϕ'_k is slower than the variation of $\phi_k(t)$. Thus, every mode $y_k(t)$ can be considered as a harmonic signal with amplitude A_k and frequency ϕ'_k for a sufficiently long time interval $[t - \delta, t + \delta]$ where $\delta \approx 2\phi/\phi'_k$ [7].

The EMD algorithm extracts these modes as described in the following four steps: (1) local maxima and minima are located in the time series data $y(t)$ and then interpolated to build an upper and a lower envelope, respectively; (2) the mean value $m(t)$ of these envelopes is determined, and the first component H_1 is built by subtracting this value from the original time series $y(t)$; (3) these two steps are repeated until the stopping criterion is satisfied, and in this case, H_1 will be equivalent to the first mode $y_1(t)$ and the residue to $y(t) - H_1$, the difference between the original time series and the first mode; and (4) steps 1–3 are repeated with the residues until all of the modes and the last residue are computed.

Mode mixing and aliasing can occur when the EMD algorithm is applied to decomposed the data [8]. A variation of the original EMD approach known as ensemble empirical mode decomposition (EEMD) [9] was proposed to overcome this: a set of trials following the EMD algorithm are performed, but mixing the original time series $y(t)$ with Gaussian white noise. Thus, the EEMD algorithm is developed in four steps: (1) Gaussian white noise is added to the original data, (2) the EMD algorithm is applied to the data mixed with white noise, (3) steps 1–2 are repeated using different white noise series, and (4) the final decomposition is obtained calculating the mean value of all trials. This way, the white noise series cancel each other, and the risk of mode mixing is significantly reduced.

On the other hand, VMD is a non-recursive signal processing method designed for decomposing non-stationary signals. The decomposition takes place by means of a constrained variational problem to calculate the bandwidth of each mode. This process consists of three steps: (1) the Hilbert transform is used to obtain the unilateral frequency spectrum for each mode, (2) an exponential tuned to the estimated center frequencies is used to shift every mode’s frequency spectrum to baseband, and (3) the bandwidth of each mode is identified using the H^1 Gaussian smoothness of the demodulated signal. As suggested in the original paper [6], the constrained variational problem can be transformed into an unconstrained problem by introducing a quadratic penalty term and Lagrangian multipliers λ as follows:

$$L(\{y_k\}, \{\omega_k\}, \lambda) = \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * y_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| y(t) - \sum_{k=1}^K y_k(t) \right\|_2^2 + \left\langle \lambda(t), y(t) - \sum_{k=1}^K y_k(t) \right\rangle \tag{3}$$

where $y(t)$ is the original time series, $\{y_k\}$ is the set of all modes, $\{\omega_k\}$ is the set of the respective center frequencies, $\delta(t)$ is the Dirac function, $*$ denotes a convolution, $\|\cdot\|_2^2$ denotes a squared L^2 -norm, and α denotes the balancing parameter of the data fidelity

constraint. Then, this unconstrained problem is solved by using a technique known as the alternate direction method of multipliers (ADMM) [10,11], which allows one to obtain the modes y_k and the center frequencies ω_k with the following expressions:

$$\hat{y}_k^{n+1}(\omega) = \frac{\hat{y}(\omega) - \sum_{i \neq k} \hat{y}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (4)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{y}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{y}_k(\omega)|^2 d\omega} \quad (5)$$

where $\hat{y}(\omega)$, $\hat{y}_k(\omega)$, and $\hat{\lambda}(\omega)$ are the Fourier transformations of $y(t)$, $y_k(t)$, and $\lambda(t)$ respectively.

Regarding the forecasting models, the basic ANN structure is known as a feedforward neural network (FFNN), which is composed of a set of three layers (input, hidden, and output layers), and the information is propagated forward to the output layers using the backpropagation algorithm [12]. Given an input $\mathbf{x} = \{x_1, \dots, x_t\}$ and a hidden layer with h neurons, the output is of the form

$$\sum_{i=1}^h \beta_i \phi(\omega_i \mathbf{x} + b_i) \quad (6)$$

where β_i represents the weights resulting from connecting the hidden and output layers (*output weights*), ω_i the weights connecting the input and hidden layers (*input weights*), b_i the biases of the neurons in the hidden layer, and ϕ the activation function.

Other types of ANNs can learn spatial and temporal features of time series data. For instance, RNNs take into consideration temporal patterns by maintaining an internal state in order to process a sequence of inputs. In order to process long-term dependencies, advanced RNN structures, such as long-short term memory (LSTM) [13], and gated recurrent units (GRU) [14] should be implemented, as basic RNNs experience vanishing and exploding gradients in this scenario [15]. On the other hand, spatial features can be extracted using CNNs. Both temporal and spatial features can be considered simultaneously by combining RNN and CNN structures [16], resulting, for instance, in CNN-GRU and CNN-LSTM models. Temporal and spatial features are also taken into consideration in temporal convolutional networks (TCN) [17], in which the convolutions are causal, meaning that the outputs are only related to the current and previous inputs.

All of the decomposition algorithms and ANN-based models can be combined to build any decomposition-based hybrid model. To make this study as comprehensive as possible, 21 models in total are considered for the simulations, resulting from the combination of the 3 decomposition algorithms (EMD, EEMD, and VMD) and the 7 forecasting models (FFNN, GRU, LSTM, CNN, CNN-GRU, CNN-LSTM, and TCN) introduced in this section.

2.2. Performance Evaluation

The performance of the models is measured using one of the most widespread metrics in the WPF literature [18], the mean absolute error (MAE):

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (7)$$

where N indicates the number of samples over the testing set, y_i the value of wind power measurements, and \hat{y}_i the value of the forecasts. To facilitate the understanding of the error measures, MAE values are normalized by the total capacity of the farm and, therefore, the normalized MAE (NMAE) is used from here onwards to report the results.

3. Data

A dataset containing historical wind power measurements from an Irish wind farm is used to carry out the simulations. Data were collected from 1 January 2017 to 30 June 2019 at a 10 min resolution. In order to benchmark the models in the most comprehensive manner, the data are divided into one-year long sets, where the first eleven months are used for training and validation and the last month as the testing set.

Figure 2 shows all of the testing sets, in which the fluctuating nature of wind power can be observed clearly from DS-1 to DS-8. This variety of wind power generation scenarios allow us to examine the performance of the models not only in terms of accuracy but in terms of robustness. Furthermore, large periods where the wind farm has been halted can be observed in the testing sets corresponding to DS-9 and DS-10.

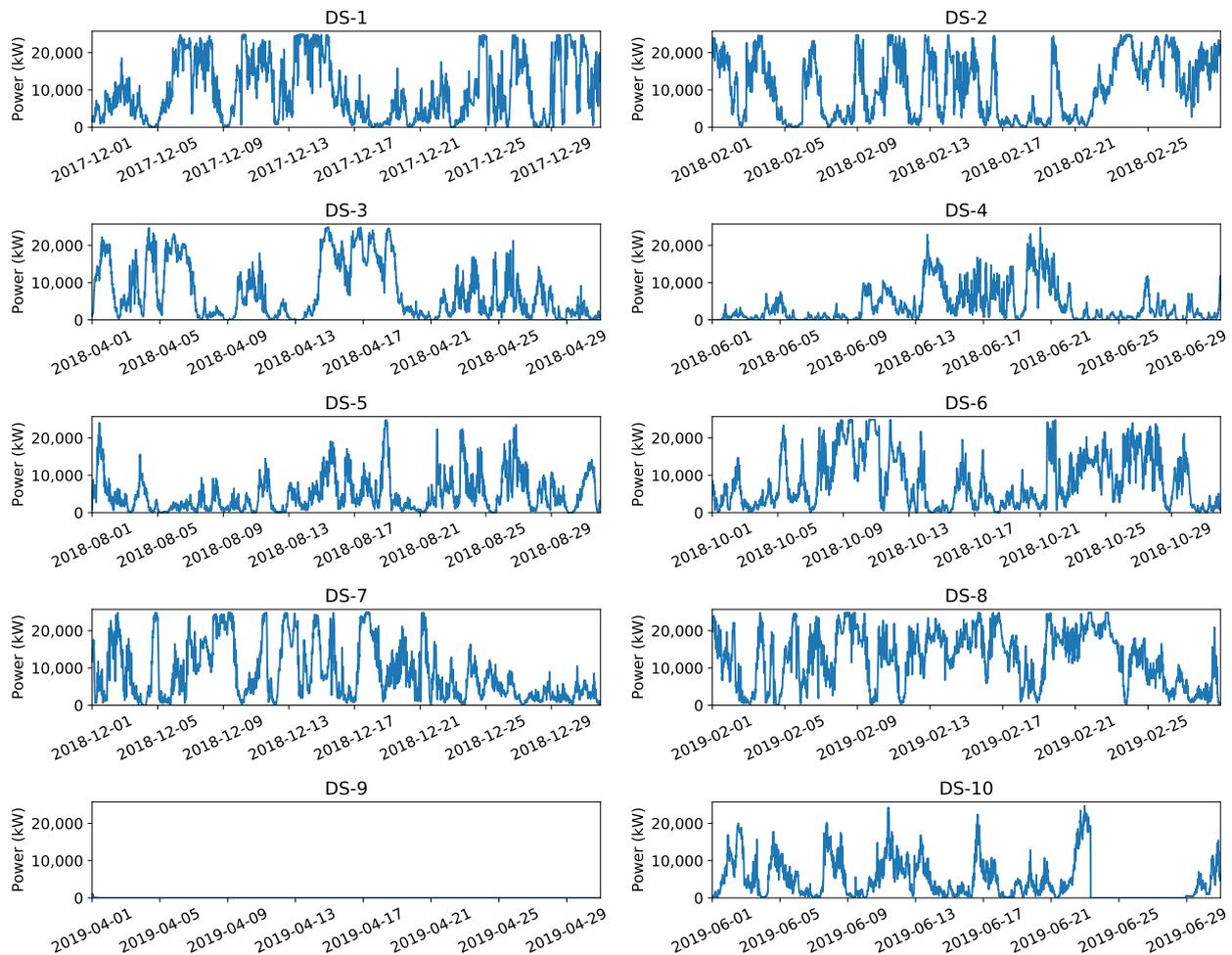


Figure 2. Testing sets considering different periods of the dataset.

4. Results

Every model was run five times for every dataset, although the results coming from training the models using the datasets DS-9 and DS-10 are omitted in this Section, as the corresponding testing sets contain large periods where the wind farm is halted, which may bias the evaluation of model performance. Thus, a total of 40 simulations were performed for all models, meaning that the models were trained 40 times, yielding different numerical results every time due to (1) the use of different subsets of data and (2) random initialization of the weights of the ANN structures, which influences the training process [19]. This way, the parameters learned by the model in the training stage vary even if the same training data are fed to the model.

Using either VMD, EMD or EEMD, data were divided into six modes, which were all trained under the same conditions. Regarding the parameters, the models were trained using a batch of size 64 for 100 epochs, using early stopping [20] to halt the training if necessary to avoid overfitting. The hidden layer of the FFNN and RNN-based models have 50 neurons in total; the CNN layers are set with 50 filters with a kernel size = 6; and the TCN layers are formed by 50 filters with dilation factors $d = 1, 2,$ and 4 and a filter size $k = 6$. The MIMO (multiple-input multiple-output) strategy [21] was implemented to output a vector with the whole sequence of forecasts, so only one model needed to be trained for all horizons. In this case, the models took the previous 72 steps as the input, representing the previous 12 h using 10 min data resolution, and produced a vector containing 36 values, which are equivalent to the next 6 h in 10 min intervals.

As only very short-term WPFs were considered in this study, the results reported correspond to 10-, 20-, and 30-min-ahead forecasts, which are equivalent to output forecasts 1, 2, and 3 steps ahead with the 10 min resolution data used in this study. Some examples of these simulations are shown in Figure 3, where 30-min-ahead forecasts are shown for DS-1, DS-2, DS-5, DS-6, and DS-8 using two of the models with better performance: the VMD-GRU and the VMD-CNN-LSTM models.

The average value of the NMAE over all the simulations is shown in Table 1. In terms of the decomposition algorithm, VMD proves to be the better than EEMD and EMD at decomposing wind power time series, as the performance using VMD is higher than that of the others in terms of accuracy, regardless of the ANN model used. Among these, the models where the temporal patterns of data are considered exhibit the best performance: an average NMAE value of 0.42 with the VMD-CNN-GRU model for 10-min-ahead forecasts; 0.59 with the VMD-GRU model for 20-min-ahead forecasts; and 0.91 with the VMD-GRU, VMD-CNN-GRU, and VMD-CNN-LSTM models for 30-min-ahead WPFs. Thus, adding the CNN layer prior to either the LSTM or GRU layer does not result in any significant improvement of performance.

Figure 4 provides additional information with respect to model performance, showing the distribution of the NMAE values over the simulations for 10-min-ahead WPFs. The combination of VMD with both GRU and LSTM structures, including the CNN-GRU and CNN-LSTM structures, appears to be the more robust among all models, as the variability is very low in terms of model performance. Furthermore, it proves the adaptability of these four models to different training and testing sets of wind power. On the other hand, EMD- and EEMD-based models not only show lower accuracy but also higher variability, which indicates a lower degree of robustness for these models.

The simulations performed in this study indicate that decomposition-based hybrid models based on the VMD algorithm for the purpose of decomposing wind power time series and RNN-based forecasting models are the most adequate for WPFs up to 30 min ahead, both in terms of forecast accuracy and robustness to different testing sets. The nature of LSTM and GRU structures is reflected in the predictions, which are able to adequately capture the temporal patterns present in the data.

Table 1. Average NMAE for very short-term forecasts.

Model	10 min	20 min	30 min
VMD-FFNN	0.77	0.97	1.13
VMD-GRU	0.43	0.59	0.91
VMD-LSTM	0.46	0.66	0.92
VMD-CNN	0.82	0.91	1.1
VMD-CNN-GRU	0.42	0.61	0.91
VMD-CNN-LSTM	0.43	0.61	0.91
VMD-TCN	0.57	0.8	1.05
EMD-FFNN	1.69	2.13	2.58
EMD-GRU	1.35	1.8	2.18
EMD-LSTM	1.31	1.71	2.1
EMD-CNN	1.62	2.02	2.31
EMD-CNN-GRU	1.3	1.72	2.08
EMD-CNN-LSTM	1.3	1.69	2.07
EMD-TCN	1.38	1.7	2.04
EEMD-FFNN	1.38	1.75	1.93
EEMD-GRU	1.23	1.56	1.71
EEMD-LSTM	1.21	1.54	1.69
EEMD-CNN	1.37	1.69	1.85
EEMD-CNN-GRU	1.22	1.54	1.7
EEMD-CNN-LSTM	1.23	1.57	1.74
EEMD-TCN	1.28	1.59	1.75

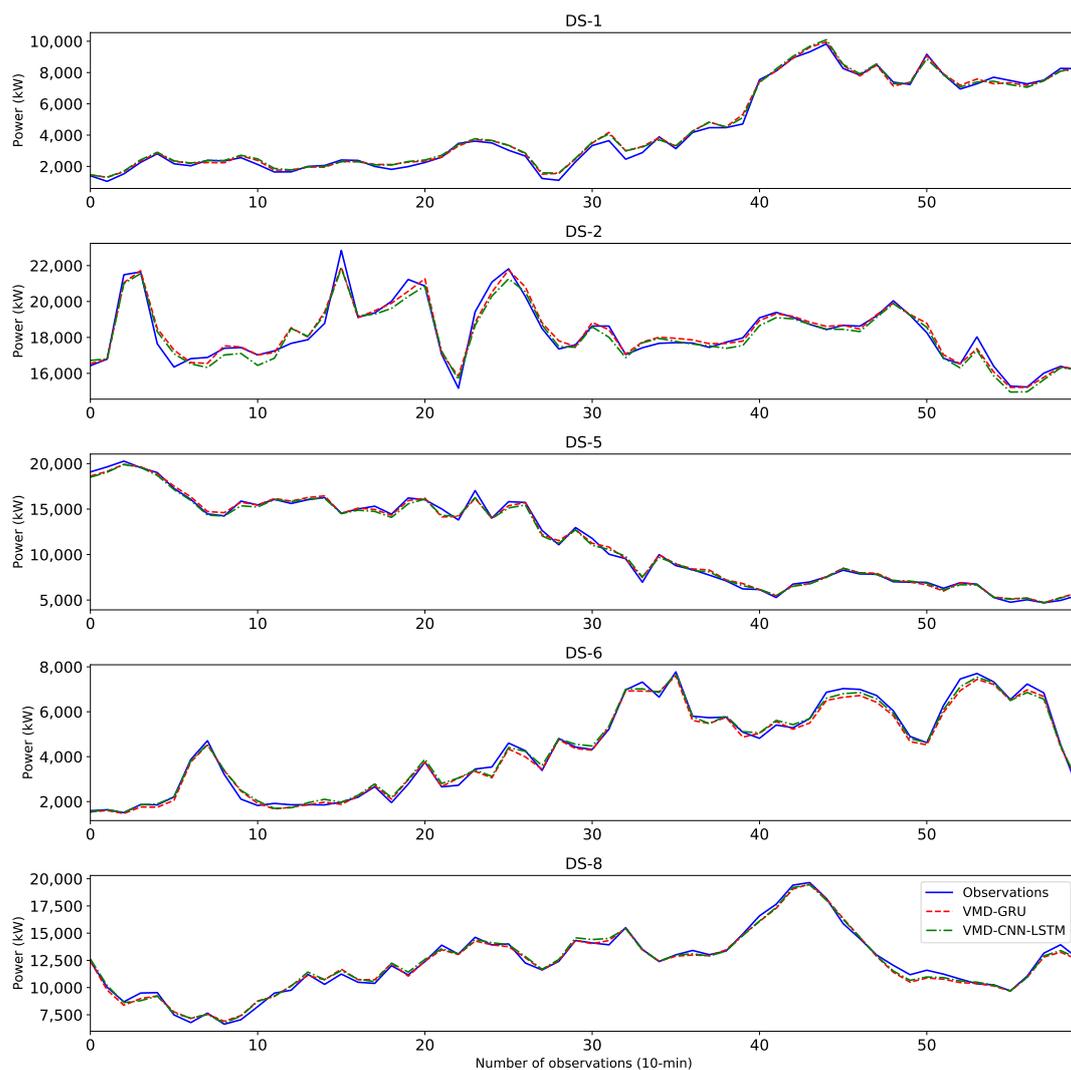


Figure 3. Examples of predictions for 30-min-ahead forecasts using the models VMD-GRU and VMD-CNN-LSTM.

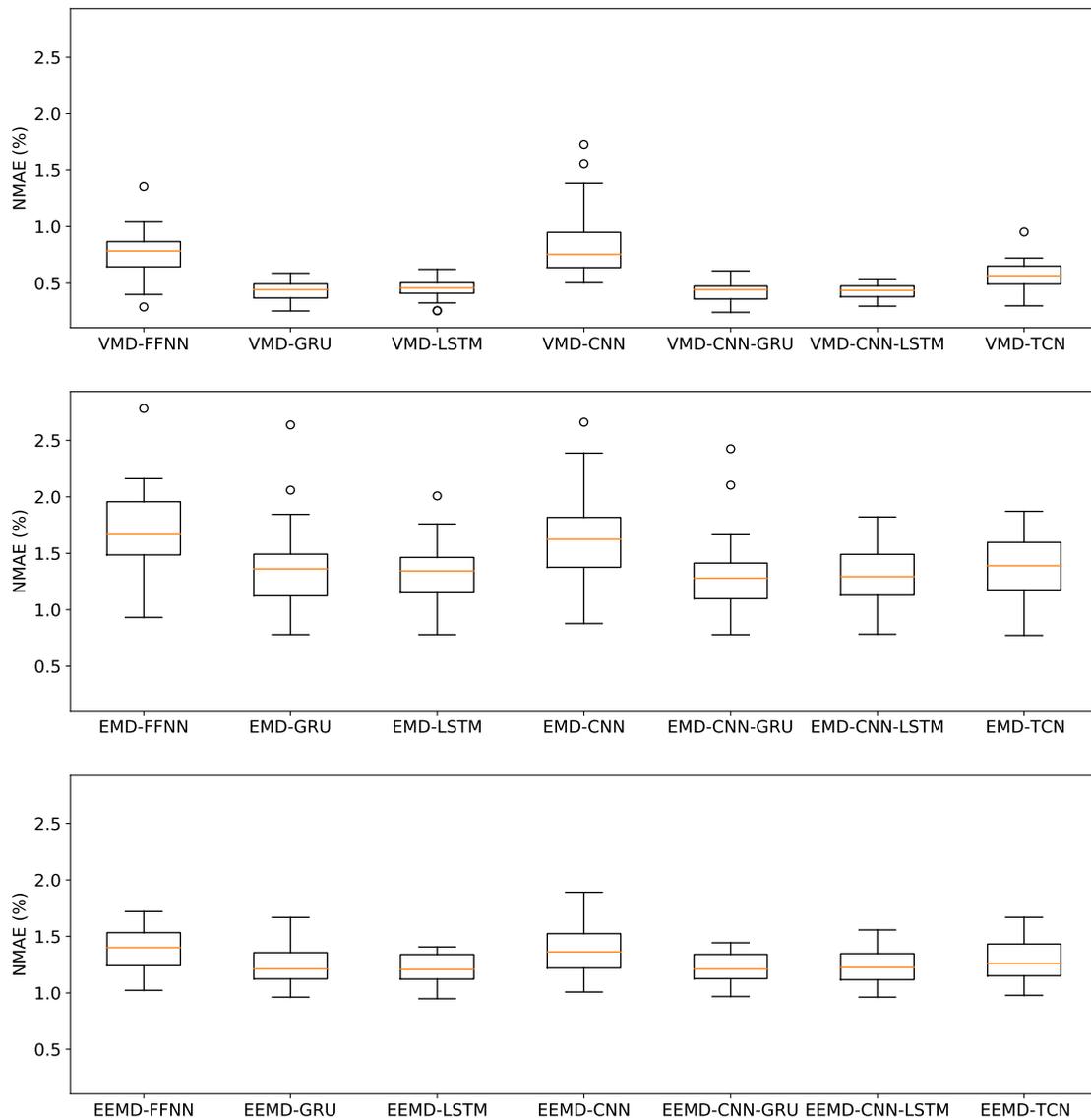


Figure 4. NMAE distribution for 10-min-ahead forecasts.

5. Conclusions

In recent times, decomposition-based hybrid models have shown promising results for very short- and short-term wind power forecasting. As the number of papers published on the topic is considerable, comparing them is a strenuous task, because the models are tested under different conditions, such as the prediction horizon, the time resolution of the data, or the amount of data used to train the model. To bring some light to this issue, this paper provides a classification of decomposition-based hybrid models for very short-term wind power forecasting, where the main state-of-the-art decomposition algorithms and the main ANN-based forecasting models are combined and benchmarked under the same conditions.

A set of simulations was performed using data from an Irish wind farm. The data were divided into several subsets to analyze the data under different training and testing conditions, as wind power time series show very high variability. As such, this study does not only identify the model performance in terms of accuracy but also their robustness to different wind power generation situations. The results indicate that using variational mode decomposition together with advanced RNN structures (LSTM and GRU) provides the most accurate and robust WPFs for very short-term horizons, showing lower average NMAE values over all the simulations and lower variability in the NMAE distribution

when compared to those of the other models. To further validate the scalability of these results, additional wind power datasets can be considered following the same experimental design shown in this paper.

Author Contributions: Conceptualization, J.M.G.S., V.P. and B.G.; methodology, J.M.G.S. and B.G.; software, J.M.G.S.; validation, J.M.G.S., V.P. and B.G.; formal analysis, J.M.G.S.; investigation, J.M.G.S., V.P. and B.G.; resources, J.M.G.S., V.P. and B.G.; data curation, J.M.G.S., V.P. and B.G.; writing—original draft preparation, J.M.G.S. and B.G.; writing—review and editing, J.M.G.S., V.P. and B.G.; visualization, J.M.G.S.; supervision, V.P. and B.G.; project administration, V.P. and B.G.; funding acquisition, V.P. and B.G. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge the funding of SEAI WindPearl Project 18/RDD/263.

Data Availability Statement: Relevant computed data are available from authors upon reasonable request.

Acknowledgments: Bidisha Ghosh would like to acknowledge the support of ENABLE (Grant number 16/SP/3804) and Connect Center (Grant number 13/RC/2077_P2). Vikram Pakrashi would like to acknowledge the support of SFI MaREI centre (Grant number RC2302_2) and UCD Energy Institute.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ADMM	Alternate direction method of multipliers
ANN	Artificial neural network
CNN	Convolutional neural network
EEMD	Ensemble empirical mode decomposition
EMD	Empirical mode decomposition
FFNN	Feedforward neural network
GRU	Gated recurrent unit
IMF	Intrinsic mode function
LSTM	Long short-term memory
MAE	Mean absolute error
NMAE	Normalized mean absolute error
RNN	Recurrent neural network
TCN	Temporal convolutional network
VMD	Variational mode decomposition
WPF	Wind power forecasting

References

1. Bessa, R.J.; Matos, M.A.; Costa, I.C.; Bremermann, L.; Franchin, I.G.; Pestana, R.; Machado, N.; Waldl, H.P.; Wichmann, C. Reserve setting and steady-state security assessment using wind power uncertainty forecast: A case study. *IEEE Trans. Sustain. Energy* **2012**, *3*, 827–836. [\[CrossRef\]](#)
2. Pinson, P.; Chevallier, C.; Kariniotakis, G.N. Trading wind generation from short-term probabilistic forecasts of wind power. *IEEE Trans. Power Syst.* **2007**, *22*, 1148–1156. [\[CrossRef\]](#)
3. Soman, S.S.; Zareipour, H.; Malik, O.; Mandal, P. A review of wind power and wind speed forecasting methods with different time horizons. In Proceedings of the North American Power Symposium 2010, Arlington, TX, USA, 26–28 September 2010; pp. 1–8.
4. Qian, Z.; Pei, Y.; Zareipour, H.; Chen, N. A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. *Appl. Energy* **2019**, *235*, 939–953. [\[CrossRef\]](#)
5. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. Ser. A* **1998**, *454*, 903–995. [\[CrossRef\]](#)
6. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2013**, *62*, 531–544. [\[CrossRef\]](#)
7. Daubechies, I.; Lu, J.; Wu, H.T. Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Appl. Comput. Harmon. Anal.* **2011**, *30*, 243–261. [\[CrossRef\]](#)
8. Mandic, D.P.; ur Rehman, N.; Wu, Z.; Huang, N.E. Empirical mode decomposition-based time-frequency analysis of multivariate signals: The power of adaptive data analysis. *IEEE Signal Process. Mag.* **2013**, *30*, 74–86. [\[CrossRef\]](#)

9. Wu, Z.; Huang, N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [[CrossRef](#)]
10. Boyd, S.; Parikh, N.; Chu, E. *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*; Now Publishers Inc.: Delft, The Netherlands, 2011.
11. Ghadimi, E.; Teixeira, A.; Shames, I.; Johansson, M. Optimal parameter selection for the alternating direction method of multipliers (ADMM): Quadratic problems. *IEEE Trans. Autom. Control* **2014**, *60*, 644–658. [[CrossRef](#)]
12. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
13. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
14. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv* **2014**, arXiv:1412.3555.
15. Pascanu, R.; Mikolov, T.; Bengio, Y. On the difficulty of training recurrent neural networks. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; pp. 1310–1318.
16. Chen, Y.; Zhang, S.; Zhang, W.; Peng, J.; Cai, Y. Multifactor spatio-temporal correlation model based on a combination of convolutional neural network and long short-term memory neural network for wind speed forecasting. *Energy Convers. Manag.* **2019**, *185*, 783–799. [[CrossRef](#)]
17. Gan, Z.; Li, C.; Zhou, J.; Tang, G. Temporal convolutional networks interval prediction model for wind speed forecasting. *Electr. Power Syst. Res.* **2021**, *191*, 106865. [[CrossRef](#)]
18. González-Sopeña, J.; Pakrashi, V.; Ghosh, B. An overview of performance evaluation metrics for short-term statistical wind power forecasting. *Renew. Sustain. Energy Rev.* **2020**, *138*, 110515. [[CrossRef](#)]
19. Pollack, J.B. Backpropagation is sensitive to initial conditions. *Complex Syst.* **1990**, *4*, 269–280.
20. Prechelt, L. Early stopping-but when? In *Neural Networks: Tricks of the Trade*; Springer: Berlin/Heidelberg, Germany, 1998; pp. 55–69.
21. Taieb, S.B.; Sorjamaa, A.; Bontempi, G. Multiple-output modeling for multi-step-ahead time series forecasting. *Neurocomputing* **2010**, *73*, 1950–1957. [[CrossRef](#)]