



Parametric Evaluation of Different ANN Architectures: Forecasting Wind Power Across Different Time Horizons

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Sewdien, V. N., Preece, R., Rueda Torres, J. L., & van der Meijden, M. A. M. M. (Accepted/In press). Parametric Evaluation of Different ANN Architectures: Forecasting Wind Power Across Different Time Horizons. In *IEEE T&D Latin America*

Published in:

IEEE T&D Latin America

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



Parametric Evaluation of Different ANN Architectures: Forecasting Wind Power Across Different Time Horizons

V.N. Sewdien
TenneT TSO B.V.
Arnhem, The Netherlands
Vinay.Sewdien@tennet.eu

R. Preece
School of Electrical &
Electronic Engineering
University of Manchester
Manchester, UK

J.L. Rueda Torres
Intelligent Electrical
Power Grids
Delft University of
Technology
Delft, The Netherlands

M.A.M.M. van der
Meijden
Intelligent Electrical
Power Grids
Delft University of
Technology
Delft, The Netherlands

Abstract— The participation of volatile wind energy resources in the generation mix of power systems is increasing. It is therefore becoming more and more crucial for system operators to accurately predict the wind power generation across different short term horizons (5 to 60 minutes ahead) in order to adequately balance the system and maintain system security. This paper presents a comprehensive assessment of the influence of different parameters in artificial neural networks, such as the amount of historic data, batch size, number of hidden layers, number of neurons per hidden layer, and the amount of training data on the short term forecast accuracy. In order to identify the parameters which are most influential with respect to forecast accuracy, a sensitivity study isolating the various factors on a one-at-a-time basis has been performed. To minimize the forecast error across the investigated forecast horizons, the developed neural networks use the feed forward back propagation algorithm. From the investigated cases it is concluded that a neural network with two hidden layers is most suitable for wind forecasting on the timeframes considered. Furthermore, with increasing forecast horizons (from 5 to 60 minutes ahead), better performance is achieved when neural networks contain increased neurons in the hidden layers and have enlarged training data sets.

Index Terms—Artificial Neural Networks, Forecasting, MIGRATE, Wind Energy

I. INTRODUCTION

With increasing penetration of wind generation it becomes essential for system operators to accurately predict future wind power injections in the system, in order to ensure reliable and affordable supply of electricity. This forecasting is done across different time horizons. Forecast models can generally be divided in two categories: statistical models and physical models. Statistical models are preferred for forecast horizons up to six hours ahead, whereas physical models perform more accurately for longer forecast horizons. Statistical models mainly use past observed data, sometimes complemented with numerical weather prediction (NWP) data. Physical models mainly use NWP data. For statistical models, artificial neural

networks (ANN) are among the top used forecasting techniques [1]. This research focuses on ANN-based statistical models for short term forecast horizons of 5, 15, 30, and 60 minutes ahead. The 5 minutes forecast horizon (FH 5) is useful for ramp forecasting, which is crucial for power systems with high penetration of wind generation [2]-[4], an example of which is given in [5]. FH 15 and FH 60 are useful for intraday markets where quarter-hourly and hourly products are traded.

The aim of this research is to investigate how the forecast accuracy across the different horizons is influenced by changes in the amount of historic data (HD), batch size (BS), number of hidden layers (HL), number of neurons per hidden layer (N_{HL}), and the amount of training data (TD).

Whereas a majority of the publications investigated the influence of the HD on the forecast accuracy, few have analyzed the impact of the HD combined with other aspects of the ANN's structure. In one study, the influence of the HD size for a single 1 hour forecast was investigated. The forecasting algorithm contained 1 hidden layer with 3 neurons, with TD 57%. It was found that the optimum size of the HD is dependent on the learning rate of the algorithm [6]. In another study the influence of HD on the forecast accuracy in terms of root mean square error for FH 30 is presented. The implemented forecasting algorithm contained 1 hidden layer, whereas the HD was varied from 3 to 8. It was concluded that the highest forecast accuracy is achieved for the ANN with HD 8 [7]. In [8] the influence of HL and HD on the forecast accuracy was investigated. It was found that a simple ANN with HD 2 and no hidden layers performed the best in terms of forecast accuracy.

The aim of these papers was to identify the ANN with the highest accuracy across one specific forecast horizon. Furthermore, the solution space considered in these papers is rather limited, as maximum two parameters were varied. Therefore there are still unresolved questions around the impact of proper tuning of the ANN's parameters on the

accuracy and how this differs across different forecast horizons.

This paper addresses these points and presents insights in the combined influence of the HD, HL, N_{HL} , TD, and BS on the forecast accuracy for forecast horizons 5, 15, 30, and 60 minutes ahead. Also, for each of these forecast horizons the impact of properly tuning the ANN's parameters is shown. With these insights it becomes possible to optimize only those parameters that have the biggest influence.

II. RESEARCH METHOD

The aim of this research is to examine the extent to which parameters and settings of the ANN influence the accuracy of wind power forecasts. The impact on forecast accuracy will be considered by observing the normalized mean absolute error. It should be noted that the focus of this work is not on minimizing the forecast error, but on observing how it is affected by variations in ANN properties. This influence is investigated for four different forecast horizons.

A. Artificial Neural Network

An ANN acts as a black box that maps inputs to outputs. In the case of wind power forecasting, it aims to map inputs such as past wind power values or NWP data to future wind power values. It will learn this input-output mapping by being trained and optimized. Full details about ANNs can be found in [9]. A brief summary of their basic form and function is provided here. Figure 1 illustrates the general architecture of an ANN.

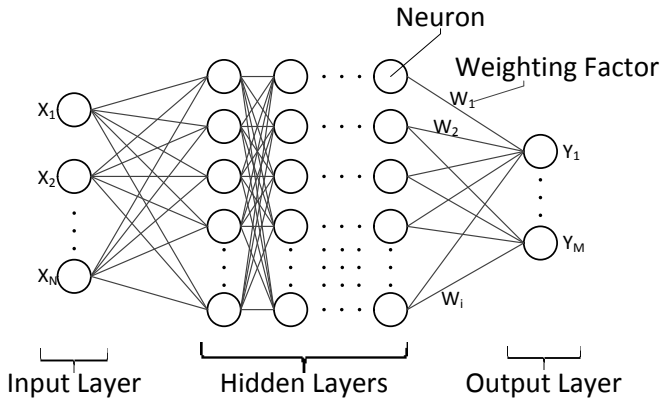


Fig. 1 Artificial Neural Network: General Architecture

The generalized ANN shown in Figure 1 consists of an input layer, one or more hidden layers, an output layer, and several synapses with their associated weighting factors. Each layer contains a number of neurons. With respect to the application of wind power forecasting, the input layer can consist of either previously observed values of the wind power generation or numerical weather prediction data (such as wind speed, pressure, and temperature). A single neuron in the input layer is assigned to each input variable. The number of neurons in the hidden layers can be chosen arbitrarily. An activation function is used to determine the weighting factors of the neurons in the last hidden layer. The dimension of the output layer is determined by the number of outputs being forecasted. The activation function implemented in the ANN in this work is the

rectifier function. The rectifier function is widely used due to its low forecast error and high sparsity [10]. Based on the objective function of the ANN's optimizer, the weighting factors are updated using the feed forward back propagation (FFBP) technique [11]. The algorithm for the FFBP technique can be decomposed in four steps. In the first step the input data is fed into the ANN, after which a forecasted value is produced. In the second step, the forecasted output is compared to the actual observed value. The error is back propagated to the output layer. In the third step, the back propagation continues to the hidden layers. In the final step, the weights are updated. This algorithm stops when the predefined number of epochs (i.e. optimization iterations) has been reached. The target of the implemented objective function is to minimize the mean absolute error.

The ANN developed for this paper is modelled in Python [12]. The parameters that were kept constant during the analysis are given in Table I.

TABLE I. FIXED PARAMETERS OF ANN

Parameter	Parameter Domain	Value
Epochs	Forecast Algorithm	100
Activation Function	Forecast Algorithm	Rectifier Linear Unit [10]
Output Layer Neurons	ANN Architecture	1
Optimizer	Forecast Algorithm	Adam [13]

The FH for which analyses were carried out are 5, 15, 30, and 60 minutes ahead. For forecast horizons up to six hours ahead better accuracies are typically achieved when using historical observed wind power generation values as inputs instead of NWP data [14]–[16]. In total 27 cases combining different permutations of the following ANN parameters were investigated for each FH (see Table V in Appendix A for full details):

- Number of inputs, i.e. the historic data set (HD): 5, 10, 20.
- Number of hidden layers (HL): 1, 2, 3.
- Number of neurons per hidden layer (N_{HL}): 100% (i.e. equal to the number of neurons in the input layer) and 50% (i.e. equal to the average of the neurons in the input and output layer).
- Size of the training data set (TD): 50% and 80% of the test data.
- Batch size (BS), i.e. amount of observations after which the weighting factors are updated: 5, 10, 20.

The characteristics of the Base Case are given in Table II. The Base Case is defined as an approximate midpoint for the parameters to be varied, however it should be noted that this is an arbitrary selection. The aim is to establish the impact of variation of these parameters on the forecast error.

TABLE II. BASE CASE VARIABLES

Variable	Value
HD	10
HL	2
N_{HL}	50 %
TD	80 %
BS	10

B. Data

The data used for this research was retrieved from the WIND Prospector Toolkit of USA's National Renewable Energy Laboratory, and belongs to a small wind park of 16 MW (Site ID 8501) [17]–[20]. Observed NWP data (wind direction, wind speed, air temperature, surface air pressure, and air density) and wind active power generation data with a resolution of 5 minutes are available for the time span 2007–2012. The statistical parametric t-test was performed successfully (i.e. rejection of the null hypothesis) on the data sets to determine if all the data belonged to the same population.

C. Forecast Error: Mean Absolute Error

When assessing the accuracy of different forecasts, root mean square error (RMSE) and mean absolute error (MAE) are the most commonly used accuracy metrics [21]. As RMSE is more sensitive to outliers (i.e. larger errors are penalized heavier) [22], the normalized MAE (nMAE) will be used as measure for the forecast accuracy. The nMAE is calculated as in Equation (1).

$$nMAE = \frac{1}{P_{MAX}} \frac{1}{n} \sum_{i=1}^n |y_i - y_{ip}|, \quad (1)$$

In Equation (1) P_{MAX} is the maximum active power generation of the wind farm, n is the number of observations, y_i is the observed wind generation for time step i , and y_{ip} is the forecasted wind generation for time step i .

As best practice dictates, for each FH a comparison is made between accuracies of the developed ANN and the persistence model [14]. In the persistence model, the predicted value for time step i is equal to the observed value at time $i-1$. New forecast should perform at least better than persistence.

For each of the 108 cases (27 cases per forecasting horizon), the ANN is trained using the data from 2007 [17]–[20]. After the training, the ANN is evaluated by calculating the nMAE for each year of data (2008 – 2012). The final nMAE reported per case in this paper is the average nMAE over the 5 years of that case. An example for case 15 for FH 5 is given in Table III. These results represent one of the lowest error cases achieved.

TABLE III. nMAE FOR CASE 15 (FH=5)

Year	nMAE
2008	2.778 %
2009	2.398 %
2010	2.622 %
2011	2.442 %
2012	2.469 %
Final nMAE	2.542 %

III. RESULTS & DISCUSSION

In Figure 2, the nMAE distribution is given for FH 5, 15, 30, and 60 minutes. The points on the graph are equally spread around a circle (the angle has no significance). Accuracies on the outer circle of 0.3 need to be disregarded: the ANN in these cases did not give any output. From Figure 2 it can be observed that the developed forecast algorithms have a low bias across all the investigated horizons. On the other hand, with

increasing forecast horizon, the general trend observed is one of an increasing variance. Low bias-low variance algorithms are preferred, as these result in algorithms with the lowest errors [23]. There is, however, always a trade-off between the bias and the variance.

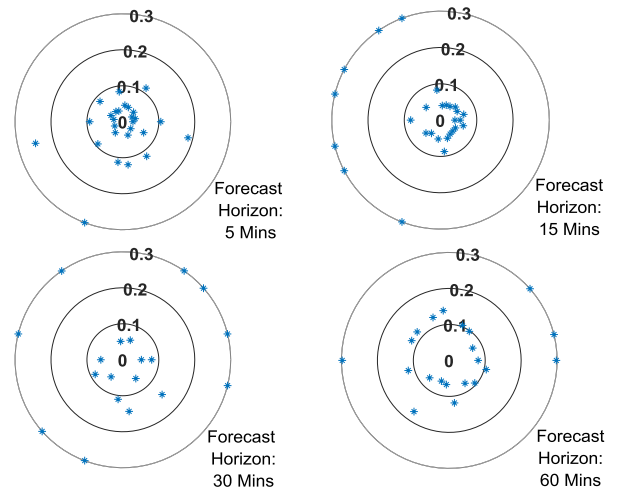


Fig. 2 nMAE Distribution for Different Forecast Horizons

In Figure 3, the nMAE of the wind power forecast is given for all the cases across the four forecasting horizons.

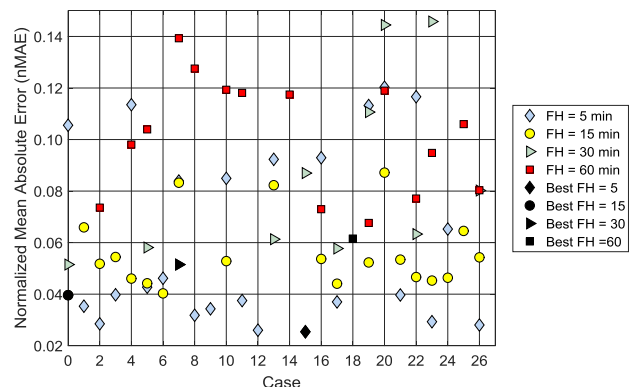


Fig. 3 Forecast Performance across Four Different Forecast Horizons

A. Forecast Horizon: 5 Minutes

For this FH, the best forecast accuracies (i.e. a lowest nMAE) is achieved for HD 20. This observation is independent on NHL and BS, as shown in Figure 4. In 8 of the 9 cases, an ANN with N_{HL} 50% outperforms an ANN with N_{HL} 100%. Only with HD 5, the ANN with HL 1 performs better. In all the other cases, the ANNs with HL 2 have a higher accuracy. The general trend is that with increasing dimension of HD, cases with TD 80% result in a slightly better performance. As the number of inputs is lower in the case with HD 5, the ANN can be trained relatively better with less data. Therefore the case with TD 50% and HD 5 achieves a higher performance. When combining the variables, the best performance is achieved for an ANN with HL 2, N_{HL} 50%, HD 5, TD 50% and BS 20. The average nMAE over the 5 years is 2.54%.

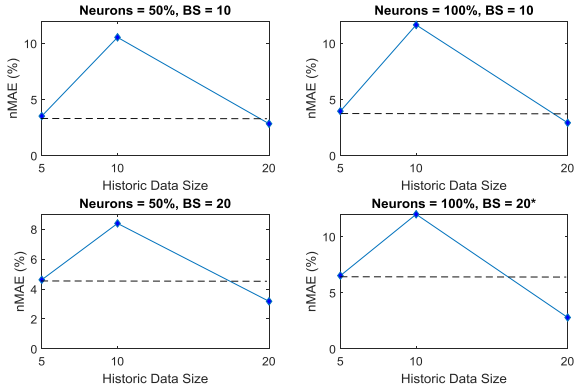


Fig. 4 Influence of Historic Data Size on nMAE for FH 5

*The actual nMAE at HD 10 is 30% (no results). For illustration purposes the nMAE is fictively fixed at 12%.

B. Forecast Horizon: 15 Minutes

The observation is that with HD 5, the highest accuracy is achieved for BS 20. With HD 20 and BS 5 the highest accuracy is achieved for the case with N_{HL} 50%. With HD 20 and BS 10 the highest accuracy is achieved for N_{HL} 100%. In 6 out of the 9 cases, an ANN with N_{HL} 50% outperforms an ANN with N_{HL} 100%. In the remaining cases, N_{HL} 100% results in a slightly lower nMAE. With HD 5, the ANN with HL 2 performs the best. With HD 10, the ANN with HL 3 layers performs best. With HD 20, no reliable result is achieved. For HD 5, the best performance is achieved for TD 80%. For HD 10 and HD 20, lowest nMAE achieved for TD 50%. When combining the variables, the best performance is achieved for an ANN with HL 2, N_{HL} 50%, HD 10, TD 80% and BS 10. The average nMAE over the 5 years is 3.96%.

C. Forecast Horizon: 30 Minutes

In 4 out of 6 cases, HD 10 resulted in a better performance. In all the cases N_{HL} 50% outperforms ANN with N_{HL} 100%. The best performance is achieved with HL 2. For HD 5 and HD 20 best performance is achieved with TD 50%. For HD 10, best performance is achieved with TD 80%. After combining various values of the parameters, the best performance is achieved for an ANN with HL 2, N_{HL} 50%, HD 10, TD 80% and BS 10. The average nMAE over the 5 years is 5.15%.

D. Forecast Horizon: 60 Minutes

In terms of the batch size, the best performance is achieved for BS 5. When HD 5, the lowest nMAE is achieved with N_{HL} 100%. For HD 10 and HD 20, N_{HL} 50% results in a higher accuracy. When varying the number of hidden layers, it is observed that an ANN with HL 3 outperforms ANNs with HL 2 or HL 1. Also, with increasing HD, a higher TD leads to increased accuracy. The overall best performance, when combining the various parameters, is achieved for an ANN with HL 2, N_{HL} 100%, HD 5, TD 80% and BS 5. The average nMAE over the 5 years is 6.15%.

The variables for the best performing ANN for each FH are given in Table IV. These models were between 43% and 52% more accurate than persistence. As expected, with increasing FH the accuracy is decreasing.

TABLE IV. VARIABLES OF BEST PERFORMING ANN

FH	Case	HL	N_{HL}	BS	TD	HD	nMAE
5	15	2	50 %	20	50 %	5	2.54 %
15	0	2	50 %	10	80 %	10	3.96 %
30	7	2	50 %	20	80 %	10	5.15 %
60	18	2	100 %	5	80 %	5	6.15 %

From the 27 investigated cases per forecasting horizon, it can be concluded that for each horizon the best performance is achieved when the ANN contains two hidden layers. Furthermore, with increasing forecast horizons, better performance is achieved when the neural networks contain relatively more neurons in the hidden layers. Up to 30 minutes ahead, 50% neurons (i.e. 3 neurons for FH 5; 5 neurons for FH 15 and FH30) results in the best accuracy. For FH 60 the best accuracy is achieved with 100% neurons (i.e. 6 neurons). The amount of training data required for optimal forecasting increased from 50% for FH 5, to 80% for the other FHs. A comparison between the best and worst performance is given in Figure 5, which clearly shows the benefits of optimal selection of ANN parameters.

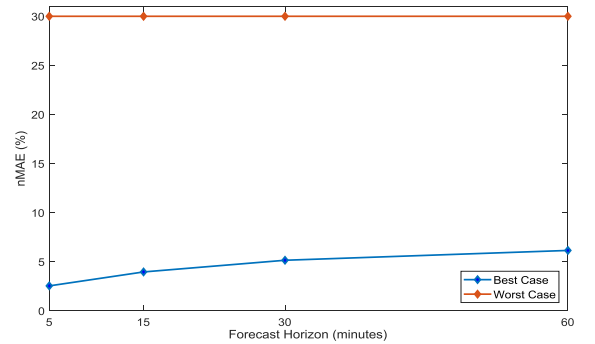


Fig. 5 Comparison between Best Case and Worst Case

IV. CONCLUSIONS

The aim of this research was to investigate the extent to which certain parameters and settings of an artificial neural network influence the accuracy of wind power forecasts across four short term forecast horizons: 5, 15, 30, and 60 minutes ahead. The results presented in this paper are based on 27 specific cases for each of the four forecast horizons. From these investigated cases it is observed that with increasing forecast horizons the variance of the forecast accuracy is increasing, whereas the bias remains low.

Furthermore, it can be concluded that the best performance is achieved when the neural network contains two hidden layers, independent of the forecast horizon. With increasing forecast horizons, better performance is achieved when neural networks contain increased neurons (100% instead of 50%) in the hidden layers and have enlarged training data sets (80% instead of 50%). The influence of the batch size and the historic data size on the forecast accuracy are dependent on the structure of the artificial neural network. When correctly choosing the ANN parameters, the nMAE decreases for FH 5 from 30% to 2.54%, for FH 15 from 30% to 3.96%, for FH 30 from 30% to 5.15% and for FH 60 from 30% to 6.15%. Compared to persistence, all models achieved at least 43% increased accuracy. As the influence of several parameters on

the forecast performance is now known, an optimization of the architecture of the artificial neural network, can be carried out with the aim to minimize the forecast error.

ACKNOWLEDGMENT



This research was carried out as part of the MIGRATE project. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 691800.

This paper reflects only the authors' views and the European Commission is not responsible for any use that may be made of the information it contains.

REFERENCES

[1] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, and Z. Yan, “A review on the forecasting of wind speed and generated power,” *Renew. Sustain. Energy Rev.*, vol. 13, no. 4, pp. 915–920, 2009.

[2] J. Zhang, M. Cui, B. M. Hodge, A. Florita, and J. Freedman, “Ramp forecasting performance from improved short-term wind power forecasting over multiple spatial and temporal scales,” *Energy*, vol. 122, pp. 528–541, 2017.

[3] A. Botterud, J. Wang, V. Miranda, and R. J. Bessa, “Wind Power Forecasting in U.S. Electricity Markets,” *Electr. J.*, vol. 23, no. 3, pp. 71–82, 2010.

[4] N. Cutler, M. Kay, H. Outhred, and I. MacGill, “High-Risk Scenarios for Wind Power Forecasting in Australia,” 2007 *Eur. Wind Energy Conf. Exhib.*, 2007.

[5] E. Ela and B. Kirby, “ERCOT Event on February 26 , 2008 : Lessons Learned,” 2008.

[6] G. Li and J. Shi, “On comparing three artificial neural networks for wind speed forecasting,” *Appl. Energy*, vol. 87, no. 7, pp. 2313–2320, 2010.

[7] M. Monfared, H. Rastegar, and H. M. Kojabadi, “A new strategy for wind speed forecasting using artificial intelligent methods,” *Renew. Energy*, vol. 34, no. 3, pp. 845–848, 2009.

[8] E. Cadenas and W. Rivera, “Short term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks,” *Renew. Energy*, vol. 34, no. 1, pp. 274–278, 2009.

[9] S. Agatonovic-Kustrin and R. Beresford, “Basic concepts of artificial neural network (ANN) modeling and its application in

pharmaceutical research,” *Journal of Ph. And Biom. Anal.*, vol. 22, pp. 717-727, 2000.

[10] X. Glorot, A. Bordes and Y. Bengio, “Deep Sparse Rectifier Neural Networks,” in *International Conference on Artificial Intelligence and Statistics*, 2011, pp. 315–323

[11] M. A. Nielsen, *Neural Networks and Deep Learning*. Determination Press, 2015.

[12] “Python Language Reference.” Python Software Foundation.

[13] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” in *International Conference on Learning Representations*, 2014, pp. 1–15.

[14] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxi, “Report: The state of the art in short-term prediction of wind power,” p. 110, 2011.

[15] G. Giebel and G. Kariniotakis, “Best Practice in Short-Term Forecasting. A Users Guide,” in *European Wind Energy Conference and Exhibition 2007*, 2007.

[16] G. Giebel and G. Kariniotakis, *Wind power forecasting—a review of the state of the art*, no. 2007. Elsevier Ltd, 2017.

[17] C. Draxl and A. C. Nrel, “A Guide to Using the WIND Toolkit Validation Code,” 2014.

[18] C. Draxl, A. Clifton, B. M. Hodge, and J. McCaa, “The Wind Integration National Dataset (WIND) Toolkit,” *Appl. Energy*, vol. 151, pp. 355–366, 2015.

[19] C. Draxl, B. Hodge, and A. Clifton, “Overview and Meteorological Validation of the Wind Integration National Dataset Toolkit,” 2015.

[20] J. King, A. Clifton, and B.-M. Hodge, “Validation of Power Output for the WIND Toolkit,” 2014.

[21] T. Chai and R. R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature,” *Geosci. Model Dev*, vol. 7, pp. 1247–1250, 2014.

[22] R. J. Hyndman and A. B. Koehler, “Another look at measures of forecast accuracy,” *Int. J. Forecast.*, vol. 22, no. 4, pp. 679–688, 2006.

[23] S. Geman, E. Bienenstock, and R. Doursat, “Neural networks and the bias/variance dilemma,” *Neural computation*, vol. 4, pp.1-58, 1992.

APPENDIX A

TABLE V. PARAMETERS OF INVESTIGATED CASES

Case	HD	HL	N _{HL}	BS	TD
0 (Base Case)	10	2	50 %	10	80 %
1	5	2	50 %	10	80 %
2	20	2	50 %	10	80 %
3	5	2	50 %	5	80 %
4	10	2	50 %	5	80 %
5	20	2	50 %	5	80 %
6	5	2	50 %	20	80 %
7	10	2	50 %	20	80 %
8	20	2	50 %	20	80 %
9	5	3	50 %	20	80 %
10	10	3	50 %	20	80 %
11	20	3	50 %	20	80 %
12	5	1	50 %	20	80 %
13	10	1	50 %	20	80 %

Case	HD	HL	N _{HL}	BS	TD
14	20	1	50 %	20	80 %
15	5	2	50 %	20	50 %
16	10	2	50 %	20	50 %
17	20	2	50 %	20	50 %
18	5	2	100 %	5	80 %
19	10	2	100 %	5	80 %
20	20	2	100 %	5	80 %
21	5	2	100 %	10	80 %
22	10	2	100 %	10	80 %
23	20	2	100 %	10	80 %
24	5	2	100 %	20	80 %
25	10	2	100 %	20	80 %
26	20	2	100 %	20	80 %