Decreasing Wind Speed Extrapolation Error via Domain-Specific Feature Extraction and Selection

Daniel Vassallo[1], Raghavendra Krishnamurthy[1, 2], and Harindra J.S. Fernando[1]

[1]University of Notre Dame, Indiana, USA
[2]Pacific Northwest National Laboratory, Washington, USA

Correspondence: Daniel Vassallo (dvassall@nd.edu)

Abstract. Model uncertainty is a significant challenge in the wind energy industry and can lead to mischaracterization of millions of dollars' worth of wind resource. Machine learning methods, notably deep artificial neural networks (ANNs), are capable of modeling turbulent and chaotic systems and offer a promising tool that can be used to produce high-accuracy wind speed forecasts and extrapolations. This paper quantifies the role of domain knowledge on ANN wind speed extrapolation accuracy using data collected using profiling lidars over three field campaigns. A series of 11 meteorological features are used as ANN inputs and the resulting output accuracy is compared with that of a simple power law extrapolation. It is found that normalized inputs, namely turbulence intensity, normalized current wind speed, and normalized previous wind speed, are the features that most reliably improve ANN accuracy, providing up to a 52% increase in extrapolation accuracy over the power law predictions. The volume of input data is also deemed important for achieving robust results. One test case is analyzed in-depth using dimensional and non-dimensional features, showing that feature normalization drastically improves network accuracy and robustness for uncommon atmospheric cases.

1 Introduction

Challenges to the prediction of microscale atmospheric flows are well documented, especially for complex terrain and forested regions (Baklanov et al., 2011; Krishnamurthy et al., 2013; Fernando et al., 2015, 2019; Sfyri et al., 2018; Yang et al., 2017; Berg et al., 2019; Wilczak et al., 2019; Pichugina et al., 2019). Poor or unfit parameterizations can lead to inaccurate flow prediction and extrapolation, producing large modeling uncertainties. Every location has unique flow features whose variability warrants a dedicated field campaign to develop and validate parameterization schemes befitting local forecasting. This process can still result in poor spatial representation of the site due to limitations in measurement technology, area covered by the field campaign, and site complexity.

Large extrapolation errors are particularly detrimental for wind farms, which rely on accurate wind speed extrapolation to estimate available wind resource and forecast output power. With the industry currently bracing for turbines up to 260m tall, vertical extrapolation accuracy has become particularly important for the next generation of wind farms. The current industry standard of 1% uncertainty per 10m vertical extrapolation (Langreder and Jogararu, 2017) must be improved in order

to increase the viability of such large-scale, powerful turbines. Additionally, it may be beneficial to move away from sensitive parameterizations towards robust machine learning techniques that consistently improve with time and data availability.

Neural networks have recently come into vogue due to the rapid increase in global computational power alongside a plethora of available data. They are particularly skilled at pattern recognition and bias correction, having been used for object recognition (Krizhevsky et al., 2012), malware detection (Saxe and Berlin, 2015), and besting classic Atari games (Mnih et al., 2013). Neural networks have even been implemented to forecast stock market prices (Moghaddam et al., 2016), showing its potential application to chaotic systems. Multiple studies have shown that neural networks perform well when tasked with wind speed forecasting on a variety of timescales (Bilgili et al., 2007; More and Deo, 2003; Chen et al., 2019). However, wind speed measurements from meteorological towers or remote sensors often must be extrapolated in space as well as time to reach the location of interest (e.g. turbine hub height), adding another layer of forecasting complexity.

In a recent study, Mohandes and Rehman (2018) found that neural networks in conjunction with lidar data can accurately extrapolate wind speeds over flat terrain using wind speeds measured below extrapolation height. However, it is unclear whether this finding holds for more complex terrain. Knowledge of meteorological conditions and site characteristics could be essential for optimal extrapolation accuracy. In the same vein, Li et al. (2019) found that adding turbulence intensity as an input greatly improves wind speed forecasting accuracy, showing that the input feature set may be highly influential for machine learning tools applied to meteorological problems. Following such developments, the present study focuses on proper extraction and selection of meteorological features, across multiple sites, for a neural network designed for vertical extrapolation of wind speed. It tries to address the following question: Is it possible to improve wind speed extrapolation accuracy in various terrains using neural networks with physics-based input features?

Section 2 provides an introduction to neural networks and the list of input features utilized. Section 3 briefly describes the campaign sites and instrumentation utilized as well as measurement uncertainty. Section 4 presents findings of the investigation, and Sect. 5 provides analysis and discussion. Concluding remarks are given in Sect. 6.

2 Model Overview

2.1 Neural Network Architecture

Artificial neural networks (ANNs) are a machine learning framework wherein a multi-layered network of nodes attempts to compute an output from a given set of inputs while eliciting (often hidden) patterns underlying a given data structure. A classic feed-forward ANN layout is given in Fig. 1. These networks mimic the inner workings of the human brain and consist of four main elements: a layer of user-defined inputs, one or more hidden layers, an output layer, and the weighted connections that adjoin any hidden layer to that before and after itself. Each layer is made up of nodes, which gather information from the previous layer, perform an activation function, and send the altered information to the next layer. ANNs with multiple hidden layers (deep neural networks) are often much better at unearthing patterns in complex, nonlinear systems. These networks learn best when supplied with large datasets and a well-selected feature set. Poor feature extraction or selection can lead the network to find a pattern that is either misleading or potentially incorrect. In other words: garbage in, garbage out.

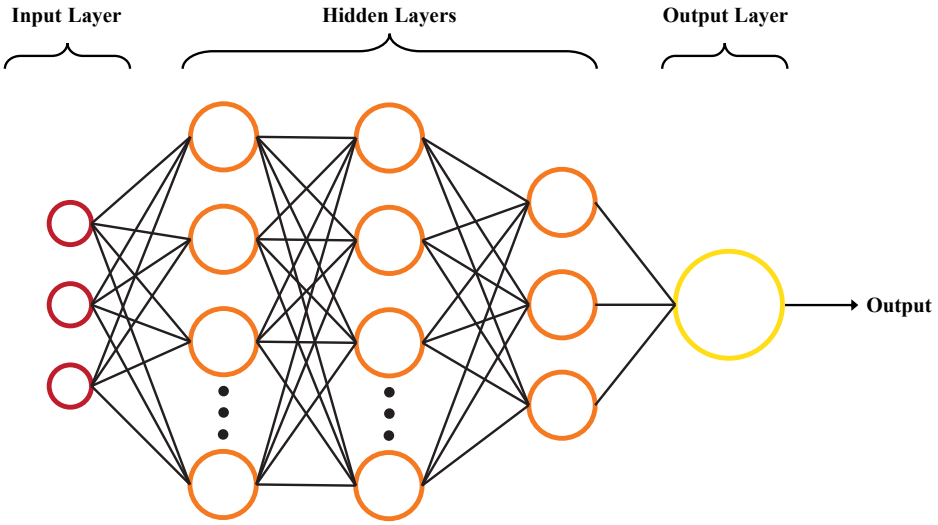

Figure 1. Artificial neural network diagram. Input nodes are in red, hidden nodes in orange, and output node in yellow. Black lines represent weighted connections between layers.

ANNs first go through a training phase where they learn the structure of a system. Batches of training data are fed into the network, which produces an output. This output is then compared to the actual output, which is known *a priori*. The network then backpropagates the error through the system via stochastic gradient descent (SGD), starting from the last layer and ending at the first. During this process, the weights between layers are altered to produce a robust network physiology. This process is repeated for as many iterations as is desired, with the network seeing all training data in each iteration.

At the end of each iteration, a set of validation data is given to the network to ensure that the network is not over-fitting the training data. At the end of training, the network is given a third set of data, known as testing data, that has been unseen by the ANN thus far. The network's performance is characterized by its prediction accuracy on the testing data, defined by a certain error or loss metric C. This study uses the mean absolute percentage error (MAPE, Eq. 1) as the loss metric due to ease of comparison with industry metrics and its universality before and after output normalization.

$$C = MAPE = \frac{1}{N}\sum_{i=1}^{N} 100 \times \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{1}$$

where N is the number of observations, y_i the observed output, and $\hat{y}_i$ the network output. The ANN used has a similar framework as Mohandes and Rehman (2018), containing four hidden layers with 30, 15, 10, and 5 nodes, respectively, descending until the final output layer that has a single node. Research on the effect of increasing the number of hidden layers shows that deeper networks are better able to approximate highly complex systems (Aggarwal, 2018). The number of hidden layers generally is a function of a) number of input and output arguments used in the ANN and b) the expected non-linearity

in the system. The true depth in an ANN is generally concluded based on several trial and error runs. Increasing the number of hidden layers in our case, however, resulted in reduced extrapolation accuracy.

There were also two dropout layers, located after the first and second hidden layers, that protected against over-fitting. The activation function in each hidden layer was the hyperbolic tangent, while the output layer had a linear activation function. The MAPE cost function was utilized as the cost function C. The Adam optimization algorithm was implemented to enhance SGD, and all trials were discontinued after no more than 1,000 iterations through the entire training dataset. All datasets were split into three distinct pieces: training data (50%), validation data (25%), and testing data (25%). In order to minimize bias, all data was randomly split before each of the 10 runs for every test case. From these 10 runs, the best, average, and standard deviation of the testing data MAPE were recorded. Tests were performed with different input features and at different heights at various site locations to confirm that bias from a given site and/or measurement height was removed. The Keras library, built on TensorFlow, was utilized to construct the ANN model (Abadi et al., 2016; Chollet et al., 2015).

2.2 Input Features

Our main hypothesis is that more informed meteorological inputs lead to lower model extrapolation error, and possibly lower error than can be achieved by existing models. To ensure that the model performs better than that achieved via simple analysis or with unadulterated inputs, we consider three base cases. The first is a power law extrapolation, a simple algorithmic representation of how wind speed varies with height,

$$WS_\alpha = WS_r\left({}^{z}/_{z_r}\right)^{\alpha} \tag{2}$$

where WS_α is the wind speed at the height of interest, WS_r the wind speed at a reference height, z the height of interest, z_r the reference height, and α the power coefficient that characterizes the shear between z and z_r. The α value was derived dynamically for each individual period (Shu et al., 2016).

The other two base cases involve using nearly raw meteorological data as input features. The second base case uses only the wind speeds below the height of interest (WS) as inputs, while the third base case uses WS, wind direction (Dir), and hour (Hr) as inputs. The hour is formatted as a cosine curve to ensure continuity between days, while the direction is formatted from $-180 \rightarrow 180$ to alleviate scaling issues.

Neural network inputs are taken at 20m intervals to a maximum of 80m below the height of interest (i.e. an output of 120m WS involves taking data from 100m, 80m, 60m, and 40m). The lowest measurement height available was 40m. Because sites (Sect. 3) had different instrumentation, the only features used are those obtained by a single profiling lidar. All lidar data is 10-minute averaged. Three normalized (i.e. non-dimensional) features are extracted, namely turbulence intensity ($\boldsymbol{TI} = {}^{\sigma_{ws}}/_{WS}$; σ_{ws} is the root mean square fluctuating wind speed), horizontal wind speeds normalized by WS 20m below the height of interest ($\boldsymbol{WS_n}$), and normalized horizontal wind speed (again from 20m below the height of interest) from the previous time period ($\boldsymbol{WS_p}$). $\boldsymbol{WS_p}$ is the only input feature utilized that extended up to the height of interest, and bold lettering on these three features indicates non-dimensionalization. Three additional features are also employed: vertical wind shear ($dudz =$

$^{\partial u}/_{\partial z}$), local terrain slope in the direction of incoming flow (ϕ), and vertical wind speed (W). The normalized input features were selected considering their robustness in inputting more accurate features (e.g., possible compensation of measurement errors in formulating non-dimensional variables) and ability of non-dimensional variables to better decompose flow structures (Barenblatt and Isaakovich, 1996). Features are used in various combinations in order to determine which provide useful information to the network and which provide unnecessary or redundant information that lead to confusion. All input features, a list of which can be seen in Appendix A, were included in a final test to show that simply throwing multitudes of data at the network yields poor results.

The subscript 1 (i.e. $\boldsymbol{WS_{p,1}}$) denotes that the input value was only taken at the height of interest, the subscript 2 (i.e. W_2) denotes that the input value was only taken at 20m below the height of interest, and the subscript 3 (i.e. $\boldsymbol{WS_{p,3}}$) denotes that the input value was taken at the height of interest and 40m below. Additionally, because a vast majority of industrial wind turbines do not produce power from exceedingly low wind speeds, all cases with $WS_1 < 3ms^{-1}$ were removed before testing.

3 Site Description and Instrumentation

Data from three international field campaigns, whose locations can be seen in Fig. 2a, were used in this study. The authors participated in each of these campaigns by deployment of instruments and data analysis. The Wind Forecasting Improvement Project 2 (WFIP2) was a multi-year field campaign focused on improving the predictability of hub-height winds for wind energy applications in complex terrain (Wilczak et al., 2019). An 18-month field campaign took place in the US Pacific Northwest from October 2015 to March 2017. Several remote sensing and in-situ sensors were located in a region with distributed commercial wind farms along the Colombia river basin. This study focuses on using vertical profiling lidar (Leosphere's Windcube V1) data from the so called Wasco site for a period of 15 months. The lidar's location can be seen as the orange marker in Fig. 2b. The surrounding terrain is reasonably complex, with neighboring wind farms to the east of the lidar. Any periods with missing data at multiple heights were ignored in the analysis. A Signal-to-Noise Ratio (SNR) and availability threshold (30%) recommended by the manufacturer is used to remove any potentially bad data.

The Coupled Air-Sea Processes for Electro-magnetic ducting Research (CASPER) field campaign was focused on measurement and modeling of the Marine Atmospheric Coastal Boundary Layer (MACBL) to better predict the interaction of EM propagation and atmospheric turbulence (Wang et al., 2018). Two field campaigns were conducted during CASPER, one near the coast of Duck, North Carolina (CASPER-East) in 2015 and another near the coast of Point-Mugu, California (CASPER-West) in 2017. This study uses data from the CASPER-West experiment. Vertical profiler data (Windcube V1) from the FLoating Instrument Platform (FLIP) was used for this study. The profiler's location can be seen as the blue marker in Fig. 2c. The data were filtered using the SNR and availability threshold recommended by the manufacturer. Datasets available at all heights were selected for this study.

The final study is the Perdigão campaign, a multinational project that took place in the Spring of 2017 aimed at improving microscale modeling for wind energy applications (Fernando et al., 2019). Conducted in the Castelo Branco region of Portugal, the campaign deployed an array of state-of-the-art sensors to measure wind flow features within and around a complex double-

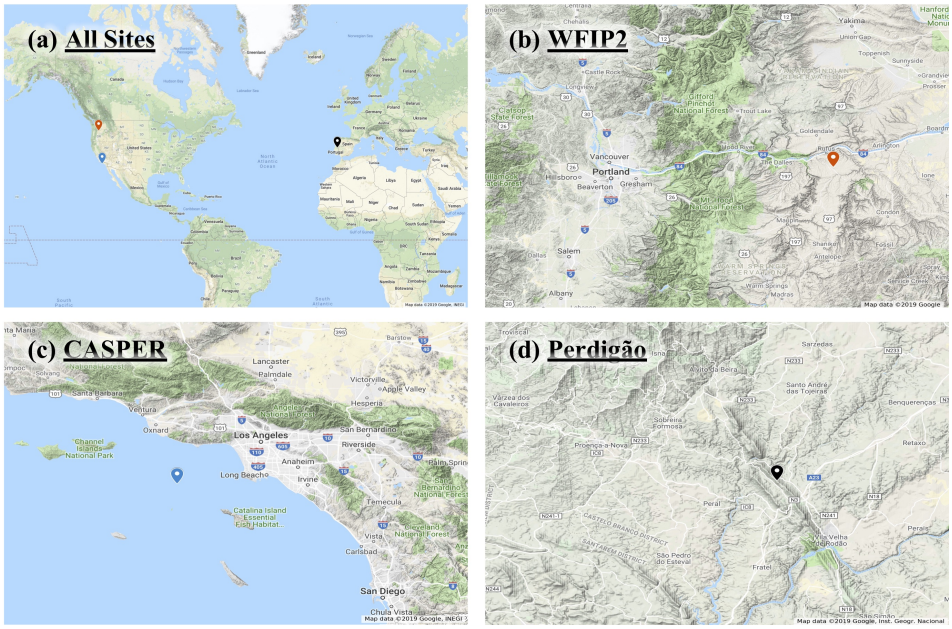

Figure 2. (a) depicts all site locations. Remaining three panels depict topography at (b) WFIP2, (c) CASPER, and (d) Perdigão. Map data ©2019 Google, INEGI and Inst. Geogr. Nacional.

ridge topography. The ridges are spaced approximately 1.4km apart with a valley in between. Both ridges rise approximately 250m above the surrounding topography, which mainly consists of rolling hills and farmland. Data were taken from a Leosphere profiling lidar, denoted by the black marker in Fig. 2d, which was located on top of the northern ridge of the Perdigão double-ridge. This particular location was selected due to the multitude of complex flow patterns seen at this location during the campaign. Profiler data available at all heights were used for this study, and any data below 30% availability over 10 minutes were ignored.

The uncertainty of the wind Doppler lidar measurements is expected to be within 2% (Lundquist et al., 2015, 2017; Giyanani et al., 2015; Kim et al., 2016; Newsom et al., 2017; Newman and Clifton, 2017).

4 Results

Fig. 3 shows for each case the best testing extrapolation accuracy at all sites. The amount of data available is also shown for reference. It is color coded, with the best accuracy in yellow and the worst in red. At first glance it is obvious that the network's accuracy is highly dependent not only on the inputs used, but also the site location. The site with the highest extrapolation accuracy is the nominally mildly complex WFIP2 site. The highly complex Perdigão site has the worst extrapolation accuracy, with the accuracy of the offshore CASPER site between the two. The best MAPE achieved for all heights, with each site below 2%, meets and often exceeds industry standards (Langreder and Jogararu, 2017).

	Perdigão			*WFIP2*		*CASPER*	
	120m (9,542 samples)	**160m (9,887 samples)**	**200m (9,720 samples)**	**100m (22,783 samples)**	**120m (22,912 samples)**	**100m (2,079 samples)**	**120m (2,050 samples)**
WS, Dir, Hr	2.95	2.93	2.36	2.28	2.77	4.3	4.92
WS	2.3	2.47	1.95	1.51	1.43	2.69	2.74
WS_α	2.45	3.59	2.17	1.74	1.44	2.18	2.91
WS_n	2.19	2.41	1.75	1.35	1.21	1.72	1.99
WS_n, TI	2.15	2.29	1.71	1.27	1.18	1.71	1.97
WS_n, $WS_{p,1}$	2.01	1.99	1.6	1.2	1.1	1.37	1.41
WS_n, TI, WS_α	2.14	2.3	1.74	1.24	1.16	1.61	1.88
WS_n, TI, dudz	2.16	2.29	1.7	1.26	1.19	1.78	1.82
WS_n, TI, ϕ	2.26	2.29	1.9	1.24	1.15		
WS_n, TI, ϕ_2	2.17	2.11	1.74	1.25	1.18		
WS_n, TI, W	2.11	2.23	1.71	1.25	1.14	1.78	1.9
WS_n, TI, $WS_{p,1}$	1.91	1.94	1.61	1.14	1.08	<u>1.29</u>	1.42
WS_n, TI, $WS_{p,3}$	1.92	<u>1.92</u>	<u>1.54</u>	<u>1.12</u>	<u>1.06</u>	1.34	<u>1.37</u>
WS_n, TI, $WS_{p,3}$, ϕ_2	1.98	1.92	1.62	1.12	1.06		
WS_n, TI, $WS_{p,3}$, WS_2	1.92	1.94	1.59	1.12	1.06	1.39	1.56
WS_n, TI, $WS_{p,3}$, W_2	<u>1.86</u>	1.94	1.56	1.12	1.08	1.39	1.41
All Inputs	2.09	2	1.73	1.28	1.29	2.16	2.25

Figure 3. Best MAPE per test case. For each height at every site, the case with the best result is underlined. Yellow is highest accuracy, red is lowest.

As Fig. 4 shows, the simple power law extrapolation typically performs considerably better than the ANN with inputs of WS, Dir, and Hr. However, network accuracy improves dramatically when inputs are simplified to just WS, showing that Dir and Hr have a negative effect on network accuracy. When WS is replaced by $\boldsymbol{WS_n}$, network accuracy again improves, providing a result 10-33% more accurate than the power law extrapolation. Both $\boldsymbol{TI}$ and $\boldsymbol{WS_{p,1}}$ were expected to have the largest effect on network accuracy and are therefore used as a second input alongside $\boldsymbol{WS_n}$. While $\boldsymbol{TI}$ improves network accuracy in all except the CASPER site, $\boldsymbol{WS_{p,1}}$ is more impactful, improving accuracy up to 52% over the power law extrapolation. Because this initial investigation is focused on determining inputs that improve extrapolation accuracy and since no other feature contains information on the flow's turbulence level, $\boldsymbol{TI}$ was primarily used as the second input for cases with three input features.

A majority of the third input features, specifically WS_α, $dudz$, ϕ, ϕ_2, and W, have negligible or negative effects on extrapolation accuracy. There are exceptions to this rule, nevertheless, as WS_α considerably improves accuracy at CASPER and ϕ_2 improves accuracy at 160m height at Perdigão. With a single exception, the best extrapolation accuracy is obtained when $\boldsymbol{WS_n}$, $\boldsymbol{TI}$, and either $\boldsymbol{WS_{p,1}}$ or $\boldsymbol{WS_{p,3}}$ are used as inputs. Adding extra input features beyond this point has, at best, a negligible impact on network extrapolation accuracy. This is best described by the final test case where all available features are forced into the network. With all inputs, the best extrapolation accuracy is up to 67% worse compared to the case's best result (100m CASPER, Fig. 3).

	Perdigão			*WFIP2*		*CASPER*	
	120m (%)	160m (%)	200m (%)	100m (%)	120m (%)	100m (%)	120m (%)
WS, Dir, Hr	-20.4	18.4	-8.8	-31.0	-92.4	-97.2	-69.1
WS	6.1	31.2	10.1	13.2	0.7	-23.4	5.8
WS_n	10.6	32.9	19.4	22.4	16.0	21.1	31.6
WS_n, TI	12.2	36.2	21.2	27.0	18.1	21.6	32.3
WS_n, $WS_{p,1}$	18.0	44.6	26.3	31.0	23.6	37.2	51.5
WS_n, TI, WS_α	12.7	35.9	19.8	28.7	19.4	26.1	35.4
WS_n, TI, dudz	11.8	36.2	21.7	27.6	17.4	18.3	37.5
WS_n, TI, ϕ	7.8	36.2	12.4	28.7	20.1		
WS_n, TI, ϕ_2	11.4	41.2	19.8	28.2	18.1		
WS_n, TI, W	13.9	37.9	21.2	28.2	20.8	18.3	34.7
WS_n, TI, $WS_{p,1}$	22.0	46.0	25.8	34.5	25.0	40.8	51.2
WS_n, TI, $WS_{p,3}$	21.6	46.5	29.0	35.6	26.4	38.5	52.9
WS_n, TI, $WS_{p,3}$, ϕ_2	19.2	46.5	25.3	35.6	26.4		
WS_n, TI, $WS_{p,3}$, WS_2	21.6	46.0	26.7	35.6	26.4	36.2	46.4
WS_n, TI, $WS_{p,3}$, W_2	24.1	46.0	28.1	35.6	25.0	36.2	51.5
All Inputs	14.7	44.3	20.3	26.4	10.4	0.9	22.7

Figure 4. ANN percent improvement over power law extrapolation. Gold denotes most improvement, while blue denotes less improvement or a decline in accuracy.

5 Discussion

A brief analysis shows that non-dimensional meteorological input features ($\boldsymbol{WS_n}$, $\boldsymbol{TI}$, and $\boldsymbol{WS_p}$) drastically improve network extrapolation accuracy, allowing it to perform much better than a simple power law extrapolation. However, this uptick in accuracy does not continue as more features are added. As can be seen in Fig. 5, using more than three input features for Perdigão actually reduced network accuracy. This is most obvious when all possible features are thrown into the network. The input noise and redundancy reduces the network's ability to find usable patterns. Excess information, much of it redundant, confuses the network.

Two tests were performed to determine whether this improvement in accuracy is derived from feature normalization. Because the network performed best at Perdigão with input features of $\boldsymbol{WS_n}$, $\boldsymbol{TI}$, and $\boldsymbol{WS_{p,3}}$, the same inputs were then given to the network, but in dimensional form (i.e. WS, σ_{ws}, and $WS_{p,3}$). The dark blue bar in Fig. 5 shows that the network performed significantly worse when given dimensional features. In fact, the network performs just as poorly with dimensional features as it does when given all the input features indiscriminately, showing that non-dimensionalization has a significant impact on network performance.

Next, the 160m Perdigão extrapolation with input features of $\boldsymbol{WS_n}$, $\boldsymbol{WS_{p,3}}$, and $\boldsymbol{TI}$ was analyzed in-depth. In order to determine the exact effects of normalization, the same inputs were then given to the network in dimensional form. The results may be seen in Fig. 6. The left column shows network outputs when given dimensional features, whereas the right column shows the results obtained using non-dimensional features (herein referred to as the dimensional and non-dimensional network, respectively). Fig. 6a and b show a comparison of true wind speed and that predicted by the network. It is immediately obvious that, upon approaching sparse sample regions, the dimensional network begins to fail, clearly under-predicting high

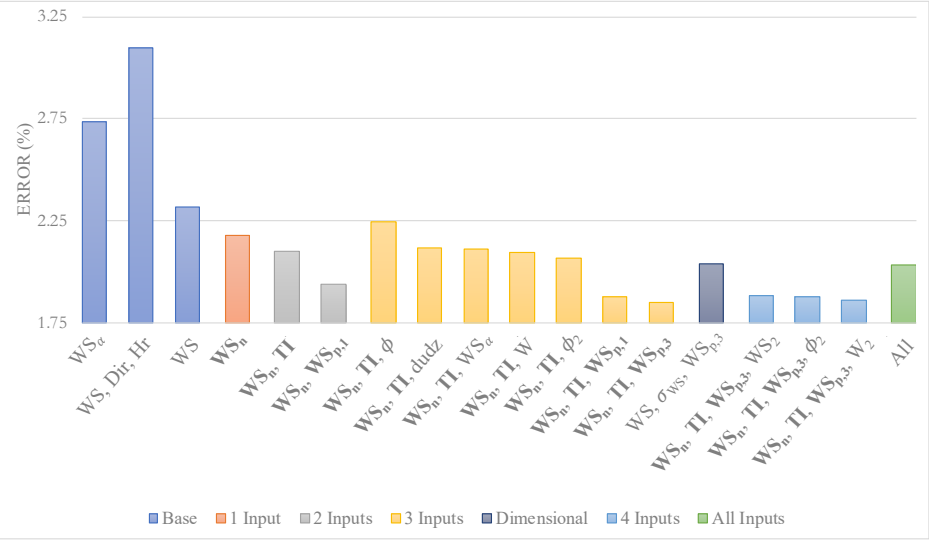

Figure 5. MAPE at Perdigão averaged over all heights. Color coding describes the number of input features used for the ANN. Dark blue bar indicates an additional test for comparison.

wind speeds. The non-dimensional network, however, does not have this problem and accurately extrapolates these extreme wind speeds.

An elementary indicator of the network's predictive power is the coefficient of determination R^2, given by

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} \tag{3}$$

where y_i and $\hat{y}_i$ have the same meanings as in Eq. 1 and $\bar{y}$ is the mean observed output. Non-dimensionalization improves R^2 from 99.3% to 99.6%. While this is a clear improvement, it does not tell the whole story. Non-dimensionalization minimizes the network's dependence on wind speed, instead shifting the focus to the amount of shear between reference and extrapolation height, which is more easily determined with the assistance of TI. Therefore, it may be expected that non-dimensionalization reduces error at high wind speeds where there is a deficiency of samples.

The decrease in error variance may be seen in Fig. 6c and d, which show the change in uncertainty η with height. For a Gaussian variable, η may be defined as

$$\eta = 100 \times \frac{\sigma_\varepsilon}{\overline{WS}} \tag{4}$$

where σ_ε is the standard deviation of error and $\overline{WS}$ is the mean wind speed. The total uncertainty, a measure of error variability, is reported in the top row of Fig. 6, but the change in η with height can be seen in the figure's middle row. At low wind speeds ($< 4ms^{-1}$) with a large sample size the dimensional network actually outperforms the non-dimensional network.

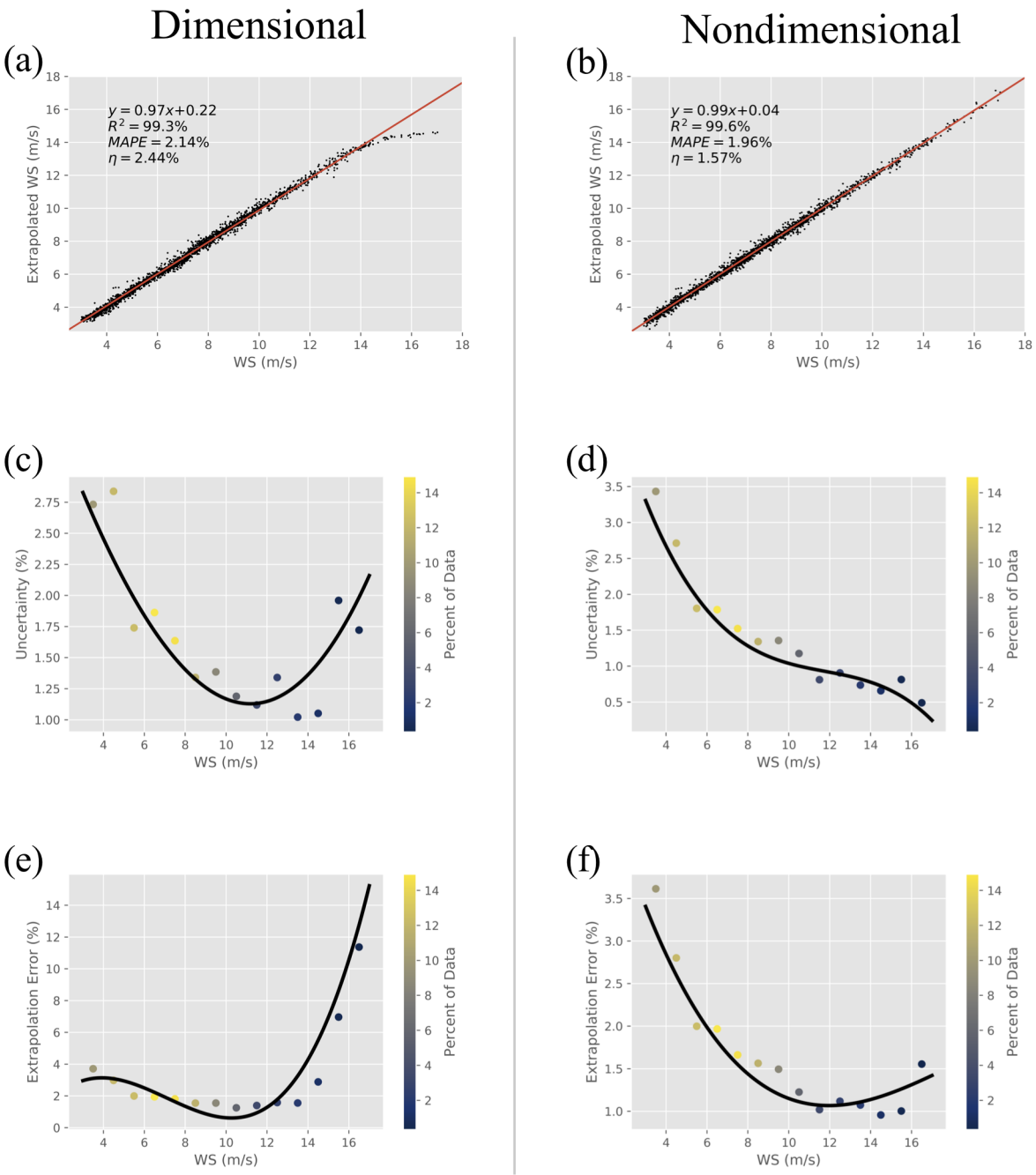

Figure 6. Comparison of network performance with dimensional (a, c, e) and non-dimensional (b, d, f) input features for the 160m extrapolation at Perdigão. Top row shows a comparison of true and extrapolated wind speed with the best-fit line, the middle row the change in uncertainty with wind speed, and the bottom row the change in extrapolation error with wind speed. The black lines indicate a spline interpolation of the data.

As wind speeds increase, both the dimensional and non-dimensional networks' uncertainties decrease at a similar rate until the sample size begins to decrease at roughly $10ms^{-1}$. At high wind speeds, the dimensional network's uncertainty begins to increase, eventually rising to almost 2% at $WS > 15ms^{-1}$. The non-dimensional network's uncertainty, meanwhile, continues to decrease as wind speed increases, eventually reaching values as low as 0.5%. Non-dimensionalization therefore decreases output variability in sparse dimensional space, producing less volatile outputs and a more robust network.

Lastly, the change in $MAPE$ with wind speed can be seen in Fig. 6e and f. As with uncertainty, the dimensional network's $MAPE$ increases dramatically with wind speed due to sample sparsity. Non-dimensionalization once again nearly eliminates this effect, as the $MAPE$ consistently decreases for $WS < 16ms^{-1}$. Whereas the uncertainty denotes error variability, $MAPE$ denotes overall prediction error. As is clear in Fig. 6a, the dimensional network has an obvious bias at high wind speeds, systematically under-predicting extrapolation wind speed. This is apparent in Fig. 6e, as $MAPE$ increases to more than 10% at extreme wind speeds. The non-dimensional network does not have this problem, again due to the fact that the network is oblivious to the dimensional wind speed value, minimizing the prediction's dependence upon total wind speed. We therefore conclude that non-dimensionalization decreases both total error and error variability in regions with a sparsity of samples by eliminating the dependence on wind speed.

CASPER is the most sensitive to the choice of input features. This may be due to two factors. First, the site may have flow patterns and dynamics for which our current list of inputs cannot account (such as the Catalina Eddy near the Californian Bight (Parish et al., 2013), and marine offshore internal boundary layers (Garratt, 1990) observed near that site). Additionally, it is likely that the amount of CASPER data available is not adequate for the network to accurately parse more complex hidden patterns. Less data could lead the network to overemphasize noisy perturbations as opposed to larger meteorological trends. It is telling that even with the small amount of data available the ANN is sometimes more than 50% more accurate than the power law extrapolation technique.

Although the best extrapolation accuracy occurs at WFIP2, the largest improvement over the power law is at CASPER and Perdigão. This unexpected finding may be due to the fact that the power law extrapolation performed so well at WFIP2 to begin with, suggesting that WFIP2 may have the simplest flow pattern of the three sites. The amount of data available did not seem to improve network performance but likely stabilized the network against noise.

We determine that of the features analyzed, the normalized input features, $\boldsymbol{WS_n}$, $\boldsymbol{WS_p}$, and $\boldsymbol{TI}$, most reliably help the ANN. Normalizing the wind speed gives the network a better idea of the general trend it needs to spot and adds more uniformity to the input samples. $\boldsymbol{TI}$ specifies the amount of turbulence and hence momentum diffusive capacity within the system (i.e. velocity gradients), a property that none of the other input features are able to directly convey. Lastly, providing the ANN with the previous period's wind speed at extrapolation height drastically improves accuracy. This is the only feature that contains information about the flow's history at extrapolation height. All three of these features are important because they give the network new insightful information about evolving aspects (dynamics) of the flow.

Some of the other input features (ϕ, W, Dir) are less impactful for extrapolation, with minor effects that are site and height dependent. Adding irrelevant inputs increases the system's noise and, unless an abundance of data is available, can cause the ANN to model coincidental or conflicting patterns. Other features ($dudz$, WS) provide redundant information. These features

typically fail to improve network accuracy, can slow the training process, and are best left out. Lastly, WS_α can act as a positive or negative influence on the network because it shows the ANN a modeled system wherein the parameter α is dependent on other parameters such as $\boldsymbol{WS_n}$, $\boldsymbol{WS_p}$, and $\boldsymbol{TI}$. If the power law model is reasonably accurate or has a clear repetitive bias, WS_α could be a useful input feature, providing the ANN with a dependable indication of wind shear. Otherwise, it adds misleading noise to the input system by thwarting the steering that $\boldsymbol{WS_n}$, $\boldsymbol{WS_p}$, and $\boldsymbol{TI}$ would provide toward an accurate extrapolation.

It is obvious that just the right amount of scaled meteorological information is necessary to achieve optimal extrapolation accuracy. It is also useful to simplify the modeled system whenever possible, provided the simplification does not remove necessary information. An example is the difference in extrapolation accuracy between WS and $\boldsymbol{WS_n}$. Normalization reduces the dimensionality of the prediction. Before normalization, the ANN has to find both a baseline wind speed and predict the vertical wind shear. After normalization, the baseline wind speed is a constant and the network only needs to calculate shear. The information lost during normalization is more than compensated by the improved model robustness and removal of some measurement inaccuracies, allowing for better generalization over regions in the input domain that would otherwise have a scarce amount of data (i.e. $WS > 14ms^{-1}$).

This is only a first step in investigating how mindful feature extraction and selection can improve ANN accuracy for meteorological predictions in wind engineering. Further improvement may be possible through the addition of other meteorological elements, particularly atmospheric stability (although we expect when inputs consist of different height levels and with specification of turbulence level, the effects of stratification is indirectly taken into account). Future studies are needed to investigate the efficacy of using non-dimensional meteorological variables to improve wind speed extrapolation and prediction. Recurrent neural networks should also be utilized to test how new combinations of meteorological features, combined with extensive knowledge of the system's history, can improve wind speed forecasting.

6 Conclusions

Model uncertainty is a vexing problem in the wind energy industry. It has been shown that the so-called 'universal' model parameters are outdated and can no longer serve their purpose in efficiently predicting and extrapolating meteorological properties (Sfyri et al., 2018; Stiperski et al., 2019). This problem can be mitigated by the use of machine learning tools which have made great strides in the past few decades. Newer and faster techniques seem to spring up every few months along with a continual increase in data processing power. ANNs have the capability to decipher turbulent, nonlinear systems and may therefore be used as a tool to assist models, although blindly using ANNs without a dynamic underpinning is vacuous. Domain knowledge can greatly assist these systems in finding the underlying trends that govern atmospheric phenomena. When properly applied, these machine learning techniques can aide in effectively calculating the rotor equivalent wind speeds for power curve measurements based on IEC 61400-12-1 Ed. 2 either from a met-mast lower than hub-height or a wind profiler with low data availability at higher heights. Additional applications in wind energy for such a technique would range from feed-forward

control of wind turbines or wind farms (Schlipf et al., 2013; Krishnamurthy, 2013; Kumar et al., 2015), prediction of yaw mis-alignment (Fleming et al., 2014), and optimizing wake steering approaches (Fleming et al., 2019).

This study investigates how feature extraction and selection can increase ANN wind speed extrapolation accuracy. Various meteorological features were combined to test their effectiveness as ANN inputs. It was found that normalizing the wind speed had a highly beneficial effect on ANN accuracy. Two other normalized features, $\boldsymbol{TI}$ and $\boldsymbol{WS_p}$, also led to increased extrapolation accuracy. The accuracy obtained by the ANN was up to 53% better than that obtained by a simple power law extrapolation. Extrapolation error was minimized to as low as 1.06% over 20m, but the use of too many network inputs actually led to a reduction in network accuracy. The 160m extrapolation at Perdigão was analyzed in depth to determine the effects of feature normalization. In addition to an improved correlation with measured wind speeds, normalization also led to a decrease in both total extrapolation error and variability, particularly at high wind speeds. The normalized input features created a robust network that yields high accuracy predictions even in rare and underrepresented cases. This shows that with sufficient data and proper feature extraction and selection, ANNs are able to improve upon current industry standards.

Future investigations are planned that will look into feature extraction and selection for wind speed predictions over a variety of timescales using a recurrent neural network, and using more reasoned non-dimensional variables to give the ANN a better perspective of atmospheric conditions. We hope that machine learning tools, combined with proper feature selection and extraction, will reduce atmospheric model uncertainty to a fraction of what it is today.

Code and data availability. Data from the Perdigão campaign may be found at https://perdigao.fe.up.pt/, the WFIP2 campaign at https://a2e.energy.gov/projects/wfip2, and the CASPER campaign at https://www.researchworkspace.com/campaign/2685070/casper-west. Processing and modeling codes were altered for every test; example codes used for this study maybe found at https://github.com/dvassall/.

Appendix A: List of Input Features

Features are listed in the order that they appear in Sect. 2.2. Normalized features are bold.

WS_α	Extrapolated wind speed based on Eq. 2 (ms^{-1})
WS	Wind speed (ms^{-1})
Dir	Wind direction ($-180 \rightarrow 180°$)
Hr	Hour of the day (cosine curve; $-1 \rightarrow 1$)
$\boldsymbol{TI}$	Turbulence intensity
$\boldsymbol{WS_n}$	Wind speed normalized by wind speed 20m below extrapolation height
$\boldsymbol{WS_p}$	Normalized wind speed from the previous time period
$dudz$	Wind shear (s^{-1})
ϕ	Terrain elevation angle from direction of incoming wind speed (°)
W	Vertical wind speed (ms^{-1})

Author contributions. Daniel Vassallo prepared the manuscript with the help of all co-authors. Data processing was performed by Daniel Vassallo, with technical assistance from Raghavendra Krishnamurthy. All authors worked equally in the review process.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was funded by the National Science Grant number AGS-1565535, Wayne and Diana Murdy Endowment at University of Notre Dame and Dean's Graduate Fellowship for Daniel Vassallo. Data collection in CASPER was funded by the US Office of Naval Research Grant N00014-17-1-3195 and WFIP2 data collection was funded by the US Department of Energy Grant DOE-WFIFP2-SUB-001.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., et al.: Tensorflow: Large-scale machine learning on heterogeneous distributed systems, arXiv preprint arXiv:1603.04467, 2016.

Aggarwal, C. C.: Neural networks and deep learning, Springer, 2018.

Baklanov, A. A., Grisogono, B., Bornstein, R., Mahrt, L., Zilitinkevich, S. S., Taylor, P., Larsen, S. E., Rotach, M. W., and Fernando, H.: The nature, theory, and modeling of atmospheric planetary boundary layers, Bulletin of the American Meteorological Society, 92, 123–128, 2011.

Barenblatt, G. I. and Isaakovich, B. G.: Scaling, self-similarity, and intermediate asymptotics: dimensional analysis and intermediate asymptotics, vol. 14, Cambridge University Press, 1996.

Berg, L. K., Liu, Y., Yang, B., Qian, Y., Olson, J., Pekour, M., Ma, P.-L., and Hou, Z.: Sensitivity of Turbine-Height Wind Speeds to Parameters in the Planetary Boundary-Layer Parametrization Used in the Weather Research and Forecasting Model: Extension to Wintertime Conditions, Boundary-layer meteorology, 170, 507–518, 2019.

Bilgili, M., Sahin, B., and Yasar, A.: Application of artificial neural networks for the wind speed prediction of target station using reference stations data, Renewable Energy, 32, 2350–2360, 2007.

Chen, Y., Zhang, S., Zhang, W., Peng, J., and 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 Conversion and Management, 185, 783–799, 2019.

Chollet, F. et al.: Keras, https://github.com/fchollet/keras, 2015.

Fernando, H., Pardyjak, E., Di Sabatino, S., Chow, F., De Wekker, S., Hoch, S., Hacker, J., Pace, J., Pratt, T., Pu, Z., et al.: The MATERHORN: Unraveling the intricacies of mountain weather, Bulletin of the American Meteorological Society, 96, 1945–1967, 2015.

Fernando, H., Mann, J., Palma, J., Lundquist, J., Barthelmie, R. J., Belo-Pereira, M., Brown, W., Chow, F., Gerz, T., Hocut, C., et al.: The Perdigao: Peering into microscale details of mountain winds, Bulletin of the American Meteorological Society, 100, 799–819, 2019.

Fleming, P., King, J., Dykes, K., Simley, E., Roadman, J., Scholbrock, A., Murphy, P., Lundquist, J. K., Moriarty, P., Fleming, K., et al.: Initial results from a field campaign of wake steering applied at a commercial wind farm–Part 1, Wind Energy Science, 4, 273–285, 2019.

Fleming, P. A., Scholbrock, A., Jehu, A., Davoust, S., Osler, E., Wright, A. D., and Clifton, A.: Field-test results using a nacelle-mounted lidar for improving wind turbine power capture by reducing yaw misalignment, in: Journal of Physics: Conference Series, vol. 524, p. 012002, IOP Publishing, 2014.

Garratt, J.: The internal boundary layer—a review, Boundary-Layer Meteorology, 50, 171–203, 1990.

Giyanani, A., Bierbooms, W., and van Bussel, G.: Lidar uncertainty and beam averaging correction, Advances in Science and Research, 12, 85–89, 2015.

Kim, D., Kim, T., Oh, G., Huh, J., and Ko, K.: A comparison of ground-based LiDAR and met mast wind measurements for wind resource assessment over various terrain conditions, Journal of Wind Engineering and Industrial Aerodynamics, 158, 109–121, 2016.

Krishnamurthy, R.: Wind farm characterization and control using coherent Doppler lidar, Ph.D. thesis, Arizona State University, 2013.

Krishnamurthy, R., Calhoun, R., Billings, B., and Doyle, J. D.: Mesoscale model evaluation with coherent Doppler lidar for wind farm assessment, Remote sensing letters, 4, 579–588, 2013.

Krizhevsky, A., Sutskever, I., and Hinton, G. E.: Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, pp. 1097–1105, 2012.

Kumar, A. A., Bossanyi, E. A., Scholbrock, A. K., Fleming, P., Boquet, M., and Krishnamurthy, R.: Field Testing of LIDAR-Assisted Feedforward Control Algorithms for Improved Speed Control and Fatigue Load Reduction on a 600-kW Wind Turbine, Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2015.

Langreder, W. and Jogararu, M.: Uncertainty of Vertical Wind Speed Extrapolation, 2017.

Li, F., Ren, G., and Lee, J.: Multi-step wind speed prediction based on turbulence intensity and hybrid deep neural networks, Energy Conversion and Management, 186, 306–322, 2019.

Lundquist, J., Churchfield, M., Lee, S., and Clifton, A.: Quantifying error of lidar and sodar Doppler beam swinging measurements of wind turbine wakes using computational fluid dynamics, Atmospheric Measurement Techniques (Online), 8, 2015.

Lundquist, J. K., Wilczak, J. M., Ashton, R., Bianco, L., Brewer, W. A., Choukulkar, A., Clifton, A., Debnath, M., Delgado, R., Friedrich, K., et al.: Assessing state-of-the-art capabilities for probing the atmospheric boundary layer: the XPIA field campaign, Bulletin of the American Meteorological Society, 98, 289–314, 2017.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M.: Playing atari with deep reinforcement learning, arXiv preprint arXiv:1312.5602, 2013.

Moghaddam, A. H., Moghaddam, M. H., and Esfandyari, M.: Stock market index prediction using artificial neural network, Journal of Economics, Finance and Administrative Science, 21, 89–93, 2016.

Mohandes, M. A. and Rehman, S.: Wind Speed Extrapolation Using Machine Learning Methods and LiDAR Measurements, IEEE Access, 6, 77 634–77 642, 2018.

More, A. and Deo, M.: Forecasting wind with neural networks, Marine structures, 16, 35–49, 2003.

Newman, J. F. and Clifton, A.: Moving Beyond 2% Uncertainty: A New Framework for Quantifying Lidar Uncertainty, Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2017.

Newsom, R. K., Brewer, W. A., Wilczak, J. M., Wolfe, D., Oncley, S., and Lundquist, J. K.: Validating precision estimates in horizontal wind measurements from a Doppler lidar, Atmospheric Measurement Techniques, 10, 2017.

Parish, T. R., Rahn, D. A., and Leon, D.: Airborne observations of a Catalina eddy, Monthly Weather Review, 141, 3300–3313, 2013.

Pichugina, Y., Banta, R., Bonin, T., Brewer, W., Choukulkar, A., McCarty, B., Baidar, S., Draxl, C., Fernando, H., Kenyon, J., et al.: Spatial Variability of Winds and HRRR-NCEP Model Error Statistics at Three Doppler-Lidar Sites in the Wind-Energy Generation Region of the Columbia River Basin, Journal of Applied Meteorology and Climatology, 2019.

Saxe, J. and Berlin, K.: Deep neural network based malware detection using two dimensional binary program features, in: 2015 10th International Conference on Malicious and Unwanted Software (MALWARE), pp. 11–20, IEEE, 2015.

Schlipf, D., Schlipf, D. J., and Kühn, M.: Nonlinear model predictive control of wind turbines using LIDAR, Wind energy, 16, 1107–1129, 2013.

Sfyri, E., Rotach, M. W., Stiperski, I., Bosveld, F. C., Lehner, M., and Obleitner, F.: Scalar-Flux Similarity in the Layer Near the Surface Over Mountainous Terrain, Boundary-layer meteorology, 169, 11–46, 2018.

Shu, Z., Li, Q., He, Y., and Chan, P.: Observations of offshore wind characteristics by Doppler-LiDAR for wind energy applications, Applied Energy, 169, 150–163, 2016.

Stiperski, I., Calaf, M., and Rotach, M. W.: Scaling, Anisotropy, and Complexity in Near-Surface Atmospheric Turbulence, Journal of Geophysical Research: Atmospheres, 124, 1428–1448, 2019.

Wang, Q., Alappattu, D. P., Billingsley, S., Blomquist, B., Burkholder, R. J., Christman, A. J., Creegan, E. D., De Paolo, T., Eleuterio, D. P., Fernando, H. J. S., et al.: CASPER: Coupled Air–Sea Processes and Electromagnetic Ducting Research, Bulletin of the American Meteorological Society, 99, 1449–1471, 2018.

Wilczak, J. M., Stoelinga, M., Berg, L. K., Sharp, J., Draxl, C., McCaffrey, K., Banta, R. M., Bianco, L., Djalalova, I., Lundquist, J. K., et al.: The Second Wind Forecast Improvement Project (WFIP2): Observational Field Campaign, Bulletin of the American Meteorological Society, 2019.

Yang, B., Qian, Y., Berg, L. K., Ma, P.-L., Wharton, S., Bulaevskaya, V., Yan, H., Hou, Z., and Shaw, W. J.: Sensitivity of turbine-height wind speeds to parameters in planetary boundary-layer and surface-layer schemes in the weather research and forecasting model, Boundary-layer meteorology, 162, 117–142, 2017.