Ref: wes-2019-58
Title: Decreasing Wind Speed Extrapolation Error via Domain-Specific Feature Extraction and Selection
Journal: Wind Energy Science

**Referee #1: Mr. Alcayaga**

We would like to once again thank Mr. Alcayaga for reading and commenting on the manuscript. His previous round of comments led to considerable improvement, and we realized a similar improvement for this round. We have addressed all comments and hope that the new version is acceptable for publication.

# P. 2 L. 26 "What is the reference to say that Machine learning techniques will be more robust (automatic parametrization) but at the same time give results as accurate as numerical models?"
**The section highlighted by the reviewer does not state that automatically parameterized machine learning methods would outperform numerical methods because (to the authors' knowledge) there has never been a study investigating whether a neural network with automatically generated hyperparameters outperforms traditional numerical models for vertical wind speed extrapolation. The goal of this statement is to explain that machine learning models may provide an alternative to numerical models and ease some of the problems which arise from numerical modeling. For example, there has been considerable research at NCAR/NOAA to replace surface layer parameterization in WRF using artificial intelligence. From many recent conference presentations, random forest and ANN models have been shown to improve upon MOST for estimating friction velocity, temperature scale and moisture scale (this was tested at multiple sites, e.g. Cabauw and Idaho, and all showed higher correlations than MOST). There is a random forest surface layer parameterization subroutine in WRF, and some results are shown in Beare et al., 2006 (but for idealized forcing).**

**The wording in P. 2 L. 5 has been changed from "Therefore, it would be beneficial to move away from such models toward robust machine learning techniques that…" to "Machine learning models may ease such concerns as they…" in order to minimize confusion.**

# P. 2 L. 32 "Reference? I supposed that this has been done before."
**Three references have been added in P. 2 L. 9. One of the references (Islam et al. 2017) refers to a study wherein genetic algorithms were utilized to automatically discover a neural network's optimal weights and biases for vertical wind speed extrapolation. While this does not directly answer the reviewer's above inquiry, the authors hope that it will be seen as tangential support for the authors' claim above that automatically parameterized machine learning models may be helpful for vertical wind speed extrapolation.**

# P. 5 L. 107 "What values were used?"
**A sentence has been added (P. 5 L. 7) which specifies the $d$ and $z_0$ values utilized for each project. A clause has also been added on the same line stating that these parameters also resulted in optimal performance for the log law extrapolation.**

# P. 5 L. 108 “How do deviations in MOST compare to the power law profile? According to this reference the larger deviations in MOST happen during non-neutral conditions. Moreover, at 120m we can be outside the surface layer (see "On the extension of the wind profile over homogeneous terrain beyond the surface boundary layer" Gryning et al, Boundary-Layer Meteorology, 2007) where a different wind profile might apply.”
**As can be seen in Table 1, the log law performs far worse than the power law for all extrapolations. However, an in-depth analysis of the log law’s performance under different atmospheric conditions is outside the scope of this work. – as such, no changes were made.**

# P. 5 L. 121 “What is the practical use of this? I suppose that if the extrapolation is necessary, there is no information available of U at this height.”
**This could be useful if, for example, the network is given the predicted vertical extrapolation from the previous time period (adding a recurrent aspect to the network’s prediction). However, for simplicity and because of data constraints, the previous period’s wind speed was directly utilized as an input. – no changes were made.**

# P. 5 L. 124 “Variables non-dimensionalized by domain or flow characteristics accomplish this and can rise similarity. Using friction velocity for example accounts for turbulence level. Not sure if you can say the same when you use U at a certain height to non-dimensionalize.”
**The reviewer’s comments are well noted. Non-dimensionalization is a standard technique in fluid dynamics which showcases trends in data which can then easily be rationally categorized/analyzed by a machine learning algorithm. We have tried to build the machine learning algorithm purely based on inputs that can be provided by a stand-alone lidar. By non-dimensionalizing U at a certain height, we (or the ANN model) can infer shear/atmospheric stability. – no changes were made.**

# P. 6 L. 134 “For non-Gaussian a median and interquartile range standarization is a common practice for machine learning.”
**Standardization is a common practice, but its primary function is to rescale all input variables such that they are on the same scale. However, the inputs in this study are already on approximately the same scale (P. 6 L. 5) and standardization is therefore redundant. This is demonstrated by the fact that standardization had no discernable impact on the results (P. 6 L. 5). Standardization is omitted in an effort to remove unnecessary complications to the study. – no changes were made.**

# P 6 L. 139 “Also 2?”
**Correct, “2” was accidentally omitted. Text on P. 6 L. 10 has been changed to “1, 2, or 3”**

# P. 6 L. 154 “It would be nice to state the -23 db and -22dB threshold, for clarity.”
**All previous references of SNR and availability thresholds have been removed, and a sentence has been added on P. 7 L. 10: “A manufacturer-recommended Signal-to-Noise Ratio (SNR) threshold (-23dB for both CASPER and Perdigão and -22dB for WFIP2) and availability threshold (30%) were used to remove any potentially bad data at all sites.”**

# P. 8 L. 193 “What extent? Is there any reference to say this?”

**A reference has been added on P. 8 L. 17 which shows the relationship between turbulence intensity and atmospheric stability. Turbulence intensity tends to increase as the atmosphere becomes unstable (i.e., convective).**

# P. 12 L. 232 "Can you give a definition of sigma_\epsilon? It is the std dev. among batches of data after random splitting and their corresponding RMSE?"
**Phrasing on P. 11 L. 7 has been changed to "standard deviation of the error (for individual predictions in the testing set)" for clarification.**

# P. 12 L. 240 "As I understand the non-dimensional input Un, you normalize by a reference wind speed, close to the extrapolation position. If you are using a dimensional height, Z, then DUn/DZ is around one order of magnitude smaller than DU/DZ. Then again, variations in this input parameter are very difficult to catch, and make it non-relevant for prediction. Might it be the irrelevance of shear that explain the improvement instead? Should this better explained by TI, and, if this is the case, maybe the non-standarization of U make difficult to catch the TI influence? Could you clarify and explain better this?"
**Non-dimensionalization does improve the prediction, as can be seen in Tables 1 and 2. Simply changing the input from U to $U_n$ consistently decreases prediction error for all cases, and we can therefore determine that it is relevant for the prediction. As is stated at the end of the paragraph (P. 11 L. 16), a majority of this improvement is due to the fact that the network must only predict the amount of shear rather than the difference between two absolute wind speeds. This alters the dimensional space and ensures that a majority of high wind speed cases (which appear as outliers when U is utilized) are seen by the model as more "common" occurrences. The shear is particularly relevant, as it is exactly that which is being predicted by the network. This is stated multiple times in Section 5 (P. 10 L. 17; P. 11 L. 15; P. 13 L. 25).**

**The improvement in prediction accuracy obtained by using TI as an input can be seen in Tables 1 and 2. As was stated in Section 2.2 (P. 6 L. 5), standardization had no discernable effect on prediction accuracy. The influence of TI is described on P. 11 L. 2 and P. 13 L. 9. It contains information about the absolute wind speed and its relation to the amount of turbulence in the atmosphere. These points are described throughout Section 5, and a brief summary of each can be found later in the section (U vs. $U_n$ on P. 13 L. 24; TI's influence on P. 13 L. 9), and therefore the authors would prefer not to restate these explanations – no changes were made.**

# P. 13 L. 295 "It is repeated several times that the non-dimensional input is one of the reason of improvement, over standarization. Is it possible to include this result? Namely, the set U, TI, normalized?"
**Standardization is the straightforward scaling of variables to reduce the range of variability, but non-dimensionalization implies that variables are selected in a way to yield variables of no-dimensions, considering nature does not know about man-made units. A rational selection of dimensionless numbers will infuse some physics into the ANN methodology, as such numbers are mostly the upshots of equations of motions, unlike variable standardization that is performed for convenience. Further explanation is given in the text "Dimensional Analysis" by G. I. Barenblatt (1987).**

**As an example of how standardization impacts prediction, we may look at the 120m extrapolation at Perdigão. The input case including (non-standardized) $U_n$ and TI obtains an average MAPE of 2.15%. However, if U and TI are standardized and used as inputs, the MAPE jumps to 4.0% (averaged over 10 runs). This is a relatively extreme case where standardization makes the prediction far less accurate. The reason for this reduction in accuracy is expected to relate to the individual rescaling of correlated variables (i.e. U inputs and output were on the same dimensional scale prior to standardization) and the fact that standardization is particularly useful when inputs and output are 1) normally distributed, and 2) cover a multitude of scales. However, none of the variables are normally distributed, meaning that all variables may not necessarily be on the same scale after standardization.**

**The authors hope that this description, which we would prefer to withhold from the manuscript as it is irrelevant for the study's findings, is acceptable for the reviewers.**

# P. 13 L. 296 "This is a bit bold. By using a simple model for extrapolation, made for surface layer conditions you can state this, but there are better models for the heights explored in this work, which is above the surface layer. See Gryning S, 2007. Rephrase it. The models don't work here because the conditions of MOST are violated, it is not a problem of up-to-date models."
**The statement made by the reviewer is equivalent to that made in the manuscript. The formulation in the reviewer's citation ("On the extension of the wind profile over homogeneous terrain beyond the surface boundary layer") relies heavily on MOST, which is why the method was tested over regions with homogeneous terrain. There are other issues with the above referenced paper, a) for very stable conditions, the model cannot be applied above several tens of meters (surface layer), b) the model does not capture very low mixing lengths observed in very stable stratification, c) it cannot model the rotation of the wind vector due to increased influence of the Coriolis force in stable conditions, above the surface layer. Hence, we expect that this model would also fail on our dataset. The modeling extrapolation methodologies typically employed in the wind energy industry are the log law and power law, and these methods are outdated for the needs of modern wind energy projects, especially in complex terrain. The citations included make similar statements about various aspects of atmospheric modeling over complex terrain. We added the clause "under various conditions" at the end of the highlighted sentence (P. 14 L. 6) because there are specific conditions wherein the current methodologies work well. However, we believe the statement made in the manuscript accurately represents the authors' (and hopefully now reviewers) intentions and therefore we have chosen to keep the wording.**

**Referee #2: Ms. Gocmen**

We would again like to thank Ms. Gocmen for reviewing and improving the manuscript with her comments and concerns. Her initial comments led to great improvement in the manuscript and we hope that this round leads to further improvements. We have attempted to address all comments and make the changes required.

# P. 5 L. 124 "See the comment next to Table 1 for a potential (alternative) explanation. - The correlation between the input features, not necessarily between the input(s) and the target, might show high dependency. That means for many of the inputs here, as you suggest several times in the text, the correlation is really high, so the inputs do depend on each other. U is indirectly used in almost all the features considered, creating bias in the optimised weight in ANN - might be one of the reasons why 'all inputs' do not necessarily perform better. Additionally, I suggest adding the clarification stated in the response to previous round of review (starts as 'P. 8 Figure "It is not realistic...") here in the main manuscript for the general audience."
**The highlighted section is kept mostly the same, with the exception of the clause ", many of which are highly correlated," which has been added in P. 5 L. 29. Two additional sentences have been added below the highlighted section to address the reviewer's concerns. These sentences (P. 5 L. 30) are similar to those from the previous round of review responses: "Including more inputs, many of which contain the same information, is not expected to improve ANN prediction accuracy because the additional information is redundant. Instead, the additional redundant or noisy inputs may decrease prediction accuracy because they reduce the ANN's ability to discern meaningful patterns contained in the pre-existing input features." We hope that this clarification is deemed acceptable and adequately expresses the points made by the reviewer.**

# P. 6 L. 134 "Fine. But standardization is not the only method for scaling. As the discussion in the previous round of reviews normalization (in NN terms) is also a common practice, similar to the wind direction scaling Dir --> (-1,1) here, I suppose. However, it is indicated that Dir is the only parameter that is scaled."
**Standardization is the industry-standard form of input and output normalization and does not improve the prediction (P. 6 L. 5; an example of the results seen when standardization is used can be seen in the response to the previous reviewer's inquiry: # P. 13 L. 295). The authors therefore assume the reviewer is referring to batch normalization or a similar process within the network itself, and therefore that is the normalization that we will address in our response. Batch normalization is becoming increasingly widespread, but it is a new enough feature that considering it an industry-wide practice would be an aggrandizement. Batch normalization could have been used, but this is one of many changes that could have been made in the modeling methodology (including convolutional layers, noise layers, or different/hybrid modeling methodology) that would potentially lead to improved prediction accuracy, but it would lead to unnecessary complications in the network architecture that are more difficult for the general audience to understand. As was stated in the previous response, this investigation is not intended to provide the absolutely optimal modeling methodology for vertical wind speed extrapolation (this would not only be tedious and deviate attention from the results, but would also be site-specific). Such fine-tuning is outside the scope of this paper. – no changes were made.**

# P. 8 Table 1 “Back to normalization/scaling discussions:
No scaling for the parameters (other than Dir) means that TI (typically in the range of 0.05 - 0.4, from offshore to complex terrain) is weighted as an input feature together with up to 100 times higher values (extrapolated wind speeds in m/s) in 'dimensional' cases. Then the performance increase is observed, when the same TI is featured together with much lower values of U_n and U_p due to non-dimensionalization. Therefore, it still suggests a NN related scaling issue for performance increase, rather than physical 'feature extraction' where every wind speed is essentially divided by a constant value. Therefore, I suggest to focus paper a little further on the framework of novelties listed on P.2 L.26, instead of potentially misleading 'performance improvements due to physically non-dimensionalized input features', which could be simply a pre-processing issue for the NN applications...”
**We believe this concern has been addressed by the case described in response to the previous reviewer’s concern about whether non-dimensionalization is in fact the reason for the network’s improvement (response to # P. 13 L. 295). Standardization of U and TI for the 120m Perdigão extrapolation led to a prediction MAPE of 4.0%, which is considerably worse than that obtained by the input case of $U_n$ and TI. This corroborates the claims that non-dimensionalization is the reason for the improvement as opposed to scaling issues when dimensional inputs are used. Therefore, we would prefer to keep the description of the findings as-is. – no changes were made.**

# P. 14 L. 304 “Given the novelties listed on page 2, I suggest to revise this sentence to better reflect the overall objective of the paper. Is it to showcase the potential of ANN for vertical extrapolation (compared to 'traditional ' models) or is it to improve the performance of ANN for this application? The former is much better suited for this paper, and potential model improvement techniques for deep learning is beyond the context of what is presented here.”
**The highlighted sentence (P. 14 L. 12) has been changed to “This study investigated the efficacy of utilizing ANNs for vertical wind speed extrapolation over a variety of terrains and evaluated how various meteorological input feature sets may influence extrapolation accuracy.” The authors believe that this is a better representation of the paper’s goals and is more consistent with the reviewer’s comments.**

# P. 14 L. 307 “too many correlated inputs, actually.”
**The statement “(many of which are highly correlated)” has been added on P. 14 L. 18, just after the highlighted section, to address this concern.**

# 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 to produce high-accuracy wind speed forecasts and extrapolations. This paper uses data collected by profiling Doppler lidars over three field campaigns to investigate the efficacy of using ANNs for wind speed vertical extrapolation in a variety of terrains, and quantifies the role of domain knowledge on ANN extrapolation accuracy. A series of 11 meteorological parameters (features) are used as ANN inputs and the resulting output accuracy is compared with that of both standard log law and power law extrapolations. It is found that extracted non-dimensional inputs, namely turbulence intensity, current wind speed, and previous wind speed, are the features that most reliably improve the ANN's accuracy, providing up to a 65% and 52% increase in extrapolation accuracy over log law and power law predictions, respectively. 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 non-dimensionalization drastically improves network accuracy and robustness for sparsely sampled 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 with variability that 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. This would likely be particularly difficult to accomplish by using numerical models. In complex terrain, these models' parameterization schemes must be tuned for each individual region in order to obtain optimal results (Stiperski et al., 2019; Bianco et al., 2019; Olson et al., 2019; Akish et al., 2019), reducing the plug-and-play effectiveness of numerical modeling. Model results must also be scaled down to the desired finer resolutions, which can result in representivity error (Dupré et al., 2020). Machine learning models may ease such concerns as they can be rapidly, economically, and automatically parameterized, require no downscaling, and consistently improve with time and data availability. When properly applied, these machine learning techniques can likely aid in effectively calculating the rotor equivalent wind speeds for power curve measurements based on IEC 61400-12-1:2017 either from a met-mast lower than hub-height or a wind profiler with low data availability at higher heights (Türkan et al., 2016; Islam et al., 2017; Mohandes and Rehman, 2018). Additional applications in wind energy for such a technique 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).

Neural networks have recently come into vogue due to the rapid increase in computational power alongside a plethora of available data. They are particularly skilled at pattern recognition and bias correction, having been used for various meteorological applications (McGovern et al., 2019; Hsieh and Tang, 1998). 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 the targeted height (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, suggesting 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. The novelty of this study is in addressing the following questions: First, is it possible to improve wind speed extrapolation accuracy under various terrain conditions using neural networks by invoking physics-based input features? Second, which atmospheric features should be selected to optimize the model's prediction capabilities?

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 Section 5 provides analysis and discussion. Concluding remarks are given in Section 6.

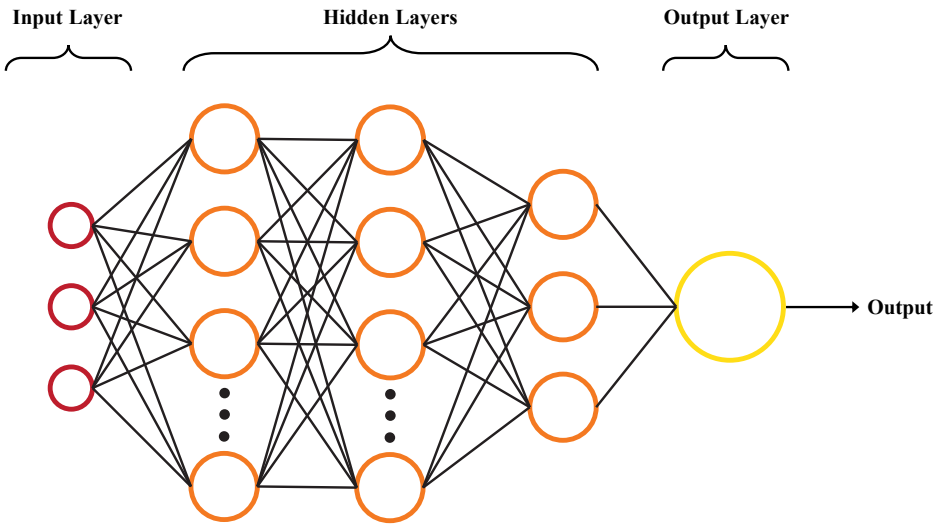


**Figure 1.** Artificial neural network diagram. Input nodes are in red, hidden nodes in orange, and output node in yellow. Black lines represent weighted connections between layers.

## 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 exploiting (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.

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 theretofore. 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 insensitivity to non-dimensionalization (Appendix A).

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

where $N$ is the number of observations, $y_i$ the observed output, and $\hat{y}_i$ the network output. The ANN used here 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 the number of input and output arguments used in the ANN as well as the perceived non-linearity in the system. The true depth of an ANN is generally concluded based on several trial and error runs. Increasing the number of hidden layers in our case, however, did not yield higher extrapolation accuracy.

There were also two dropout layers, located after the first and second hidden layers, that protect 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 (Kingma and Ba, 2014) 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 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, especially in complex terrain. All meteorological inputs utilized in this study are listed alongside their respective definitions in Appendix B. To ensure that the model performs better than that achieved via simple analysis or with unadulterated inputs, we consider four base cases. The first is a power law extrapolation, a simple algorithmic representation of how wind speed varies with height,

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

where $U_\alpha$ is the streamwise wind speed at the height of interest, $U_r$ the streamwise wind speed at a reference height, $z$ the height of interest, $z_r$ the reference height, and $\alpha$ a power law coefficient that characterizes the shear between $z$ and $z_r$. The $\alpha$ value was derived dynamically for each individual period (Shu et al., 2016). The second base case is the log law extrapolation under neutral conditions. The formulation of the log law when the wind speed is known at a reference height can be given as

$$U_L = U_r \frac{ln\left({}^{(z-d)}/_{z_0}\right)}{ln\left({}^{(z_r-d)}/_{z_0}\right)} \tag{3}$$

where $U_L$ is the wind speed at extrapolation height, $d$ the zero-plane displacement, and $z_0$ the roughness length. Both $d$ and $z_0$ are determined based on optimal testing results and local topographic information (Holmes, 2018). $d$ and $z_0$ are found to be 6m and 0.1m, respectively, for the Perdigão campaign, 0m and 0.001m, respectively, for the CASPER campaign, and 0m and 0.01m, respectively, for the WFIP2 campaign (which are the campaigns analyzed in this study, to be discussed in Section 3). The log law extrapolation (and more generally Monin-Obukhov similarity theory) is expected to perform poorly for the complex terrain sites due to the lack of stationarity and horizontal homogeneity (Fernando et al., 2015).

The other two base cases involve using nearly raw meteorological data as input features. The third base case uses only the streamwise wind speeds ($U$) below the height of interest as inputs, while the fourth base case uses $U$, 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 $-1 \rightarrow 1$ to alleviate scaling issues.

Neural network inputs are taken at 20m intervals to a maximum of 80m below the height of interest (e.g., for an output at 120m, data from 100m, 80m, 60m, and 40m are used). The lowest measurement height available was 40m. Because sites (Section 3) had different instrumentation, the only features used are those obtained by a single profiling lidar. All lidar data are 10-minute averaged. Three non-dimensional features are extracted from the lidar data, namely turbulence intensity ($\boldsymbol{TI} = {}^{\sigma_U}/_{U}$; $\sigma_U$ is the standard deviation of the wind speed), non-dimensional streamwise wind speeds ($\boldsymbol{U_n}$; non-dimensionalized by $U$ 20m below the height of interest), and non-dimensional streamwise wind speed from the previous time period ($\boldsymbol{U_p}$). $\boldsymbol{U_p}$ is the only input feature utilized that extended up to the height of interest (i.e. we assume that the previous period's wind speed at the extrapolation height is known), and bold lettering on these three features indicates that they are non-dimensional quantities. Three additional features are also extracted: vertical wind shear ($dudz = {}^{\partial U}/_{\partial z}$), local terrain slope in the direction of incoming flow ($\phi$), and vertical wind speed ($W$). The non-dimensional 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 represent 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, many of which are highly correlated, are included in a final test to show that simply throwing multitudes of data at the network yields poor results. Including more inputs, many of which contain the same information, is not expected to improve ANN prediction accuracy because much of the additional information is redundant. Instead, the additional redundant or noisy inputs may decrease prediction accuracy because they reduce the ANN's ability to discern meaningful patterns contained in the pre-existing input features.

It is typical industry practice to normalize (i.e. standardize) input variables, wherein an input variable $x$ is scaled to $\hat{x}$ via

$$\hat{x} = \frac{x - \mu}{\sigma} \tag{4}$$

where $\mu$ is the variable's mean and $\sigma$ is the variable's standard deviation (Aggarwal, 2018). This technique is particularly useful when input variables have Gaussian distributions and cover multiple scales. However, none of the variables in our study had such a distribution, and many inputs already have similar scaling. Testing showed that standardization had no discernible impact on network performance (not shown), and therefore the input features were kept in their unaltered state. The non-dimensionalization performed followed typical fluid dynamical practices (Barenblatt and Isaakovich, 1996).

The subscript 1 (e.g., $\boldsymbol{U_{p,1}}$) denotes that the input value was only taken at the height of interest, subscript 2 (e.g., $W_2$) denotes that the input value was taken at 20m below the height of interest, and subscript 3 (e.g., $\boldsymbol{U_{p,3}}$) denotes that the input value was taken at the height of interest and 40m below. Input variables without a subscript 1, 2, or 3 were taken from all four heights below the extrapolation height. Additionally, because a vast majority of industrial wind turbines do not produce power at exceedingly low wind speeds, all cases with streamwise velocity 20m below the extrapolation height $(U_1) < 3ms^{-1}$ were removed before testing. The highest wind speed value recorded at any site was less that 23 $ms^{-1}$, below the standard cut-off limit of 25 $ms^{-1}$ (Markou and Larsen, 2009).

## 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 Columbia river basin. This study focuses on using vertical profiling lidar (Leosphere's Windcube V1) data collected by the University of Colorado at Boulder from the so-called Wasco Site for a period of 15 months (Bodini et al., 2019; Lundquist, 2017). The lidar's location can be seen as the orange marker in Fig. 2b. The surrounding terrain is complex (although nominally less so than that at Perdigão to be described below), with neighboring wind farms to the east of the lidar. Any periods with missing data at multiple heights were ignored in the analysis.

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), collected over approximately a month, was used for this study. The profiler's location can be seen as the blue marker in Fig. 2c. Datasets available at all heights were selected for this study.

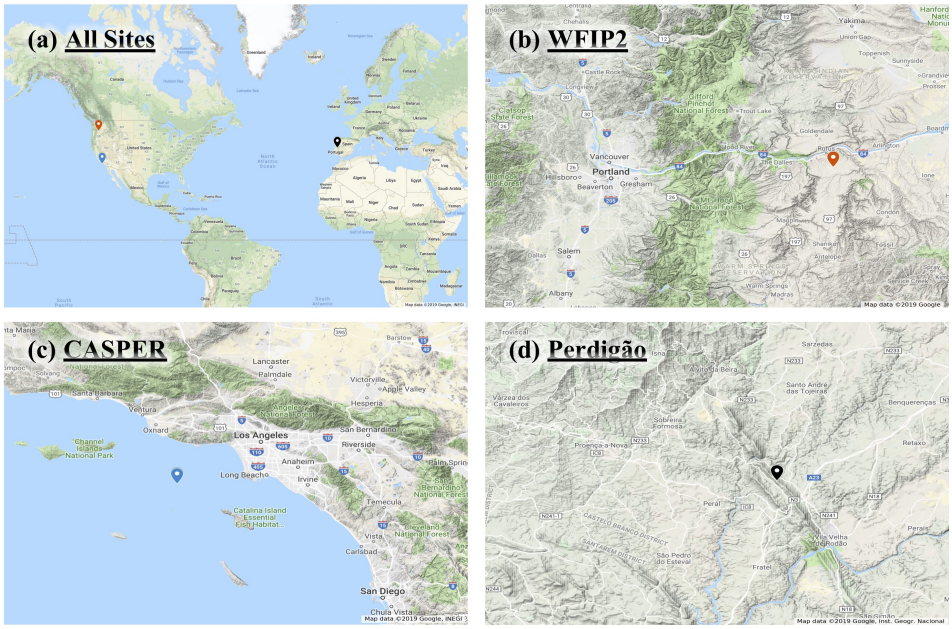


**Figure 2.** (a) depicts all site locations. Remaining three panels depict topography at (b) WFIP2, (c) CASPER, and (d) Perdigão. Map data ©2019 Google, INEGI and Inst. Geogr. Nacional.

The final study is the Perdigão campaign, a multinational project that took place in the Spring and Summer 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-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. Over four months of 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. A meteorological tower was located adjacent to the lidar, but it only rose to 100m above ground level, below all extrapolation heights. Profiler data available at all heights were used for this study. A manufacturer-recommended Signal-to-Noise Ratio (SNR) threshold (–23dB for both CASPER and Perdigão and –22dB for WFIP2) and availability threshold (30%) were used to remove any potentially bad data at all sites.

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). Owing to a lack of secondary measurements at the locations and heights of interest, all lidar measurements are treated as true.

| | *Perdigão* | | | *WFIP2* | | *CASPER* | |
|---|---|---|---|---|---|---|---|
| | 120m (4,772 samples) | 160m (4,944 samples) | 200m (4,860 samples) | 100m (11,392 samples) | 120m (11,456 samples) | 100m (1,040 samples) | 120m (1,026 samples) |
| U, Dir, Hr | 2.30 | 2.48 | 1.94 | 1.57 | 1.41 | 2.59 | 2.81 |
| U | 2.30 | 2.47 | 1.95 | 1.51 | 1.43 | 2.69 | 2.74 |
| $U_\alpha$ | 2.45 | 3.59 | 2.17 | 1.74 | 1.44 | 2.18 | 2.91 |
| $U_L$ | 4.39 | 4.69 | 3.82 | 3.06 | 2.71 | 3.70 | 3.90 |
| $U_n$ | 2.19 | 2.41 | 1.75 | 1.35 | 1.21 | 1.72 | 1.99 |
| $U_n$, TI | 2.15 | 2.29 | 1.71 | 1.27 | 1.18 | 1.71 | 1.97 |
| $U_n$, $U_{p,1}$ | 2.01 | 1.99 | 1.60 | 1.20 | 1.10 | 1.37 | 1.41 |
| $U_n$, TI, $U_\alpha$ | 2.14 | 2.30 | 1.74 | 1.24 | 1.16 | 1.61 | 1.88 |
| $U_n$, TI, dudz | 2.16 | 2.29 | 1.70 | 1.26 | 1.19 | 1.78 | 1.82 |
| $U_n$, TI, $\phi$ | 2.26 | 2.29 | 1.90 | 1.24 | 1.15 | | |
| $U_n$, TI, $\phi_2$ | 2.17 | 2.11 | 1.74 | 1.25 | 1.18 | | |
| $U_n$, TI, W | 2.11 | 2.23 | 1.71 | 1.25 | 1.14 | 1.78 | 1.90 |
| $U_n$, TI, $U_{p,1}$ | 1.91 | 1.94 | 1.61 | 1.14 | 1.08 | 1.29 | 1.42 |
| $U_n$, TI, $U_{p,3}$ | 1.92 | 1.92 | 1.54 | 1.12 | 1.06 | 1.34 | 1.37 |
| $U_n$, TI, $U_{p,3}$, $\phi_2$ | 1.98 | 1.92 | 1.62 | 1.12 | 1.06 | | |
| $U_n$, TI, $U_{p,3}$, $U_2$ | 1.92 | 1.94 | 1.59 | 1.12 | 1.06 | 1.39 | 1.56 |
| $U_n$, TI, $U_{p,3}$, $W_2$ | 1.86 | 1.94 | 1.56 | 1.12 | 1.08 | 1.39 | 1.41 |
| All Inputs | 2.09 | 2.00 | 1.73 | 1.28 | 1.29 | 2.16 | 2.25 |

**Table 1.** 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. The total number of validation and testing samples is shown for reference.

## 4 Results

Table 1 shows for each case the best testing extrapolation accuracy at all sites. The total number of (randomly split) validation and testing samples for each case is also shown for reference. The table 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 and data availability. The site with the highest extrapolation accuracy is the nominally mildly complex WFIP2 site, which also has the most robust dataset. 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 (underlined), with each site below 2%, meets and often exceeds industry standards (Langreder and Jogararu, 2017).

The power law performed better than the log law and was therefore used as a baseline for comparison in Table 2. As this table shows, the two ANN baseline cases (one utilizing $U$, $Dir$, and $Hr$, as well one only utilizing $U$; first two rows of Tables 1 and 2) performed almost equally well and showed a slight improvement over the power law extrapolation. However, there is no clear distinction between the results of the two cases and therefore $Dir$ and $Hr$ can be presumed to have no effect on prediction accuracy. When $U$ is replaced by $\boldsymbol{U_n}$, the network accuracy again improves, providing a result 10-33% more accurate than the power law extrapolation. $\boldsymbol{TI}$ and $\boldsymbol{U_{p,1}}$ are the most beneficial secondary input features when used alongside $\boldsymbol{U_n}$. While $\boldsymbol{TI}$ improves network accuracy in all except at the CASPER site, $\boldsymbol{U_{p,1}}$ is more impactful, improving accuracy up to 52% over the power law extrapolation. $\boldsymbol{TI}$ was chosen as the second input for cases with three input features because it is the most beneficial feature that includes information about the flow's turbulence levels and to some extent the atmospheric stability (Wharton and

| | *Perdigão* | | | *WFIP2* | | *CASPER* | |
|---|---|---|---|---|---|---|---|
| | 120m (%) | 160m (%) | 200m (%) | 100m (%) | 120m (%) | 100m (%) | 120m (%) |
| U, Dir, Hr | 6.1 | 30.9 | 10.6 | 9.8 | 2.1 | -18.8 | 3.4 |
| U | 6.1 | 31.2 | 10.1 | 13.2 | 0.7 | -23.4 | 5.8 |
| $U_n$ | 10.6 | 32.9 | 19.4 | 22.4 | 16.0 | 21.1 | 31.6 |
| $U_n$, TI | 12.2 | 36.2 | 21.2 | 27.0 | 18.1 | 21.6 | 32.3 |
| $U_n$, $U_{p,1}$ | 18.0 | 44.6 | 26.3 | 31.0 | 23.6 | 37.2 | 51.5 |
| $U_n$, TI, $U_\alpha$ | 12.7 | 35.9 | 19.8 | 28.7 | 19.4 | 26.1 | 35.4 |
| $U_n$, TI, dudz | 11.8 | 36.2 | 21.7 | 27.6 | 17.4 | 18.3 | 37.5 |
| $U_n$, TI, $\phi$ | 7.8 | 36.2 | 12.4 | 28.7 | 20.1 | | |
| $U_n$, TI, $\phi_2$ | 11.4 | 41.2 | 19.8 | 28.2 | 18.1 | | |
| $U_n$, TI, W | 13.9 | 37.9 | 21.2 | 28.2 | 20.8 | 18.3 | 34.7 |
| $U_n$, TI, $U_{p,1}$ | 22.0 | 46.0 | 25.8 | 34.5 | 25.0 | 40.8 | 51.2 |
| $U_n$, TI, $U_{p,3}$ | 21.6 | 46.5 | 29.0 | 35.6 | 26.4 | 38.5 | 52.9 |
| $U_n$, TI, $U_{p,3}$, $\phi_2$ | 19.2 | 46.5 | 25.3 | 35.6 | 26.4 | | |
| $U_n$, TI, $U_{p,3}$, $U_2$ | 21.6 | 46.0 | 26.7 | 35.6 | 26.4 | 36.2 | 46.4 |
| $U_n$, TI, $U_{p,3}$, $W_2$ | 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 |

**Table 2.** ANN percent improvement over power law extrapolation. Gold denotes most improvement, while blue denotes less improvement or a decline in accuracy.

Lundquist, 2012), information that is expected to be highly influential in determining the flow, particularly at the complex terrain sites.

A majority of the third input features, specifically $U_\alpha$, $dudz$, $\phi$, $\phi_2$, and $W$, have negligible or negative effects on extrapolation accuracy. There are exceptions to this rule, nevertheless, as $U_\alpha$ considerably improves accuracy for CASPER and $\phi_2$ improves accuracy at 160m height for Perdigão. With a single exception, the best extrapolation accuracy is obtained when $\boldsymbol{U_n}$, $\boldsymbol{TI}$, and either $\boldsymbol{U_{p,1}}$ or $\boldsymbol{U_{p,3}}$ are used as inputs. Adding extra input features beyond this point has, at best, 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 input case that obtains the best result (100m CASPER, Table 1).

## 5 Discussion

A brief analysis shows that extracted non-dimensional meteorological input features ($\boldsymbol{U_n}$, $\boldsymbol{TI}$, and $\boldsymbol{U_p}$) drastically improve the network's extrapolation accuracy, allowing it to perform much better than conventional log law and power law extrapolations. However, this uptick in accuracy does not continue as more features are added. As can be seen in Fig. 3, 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.

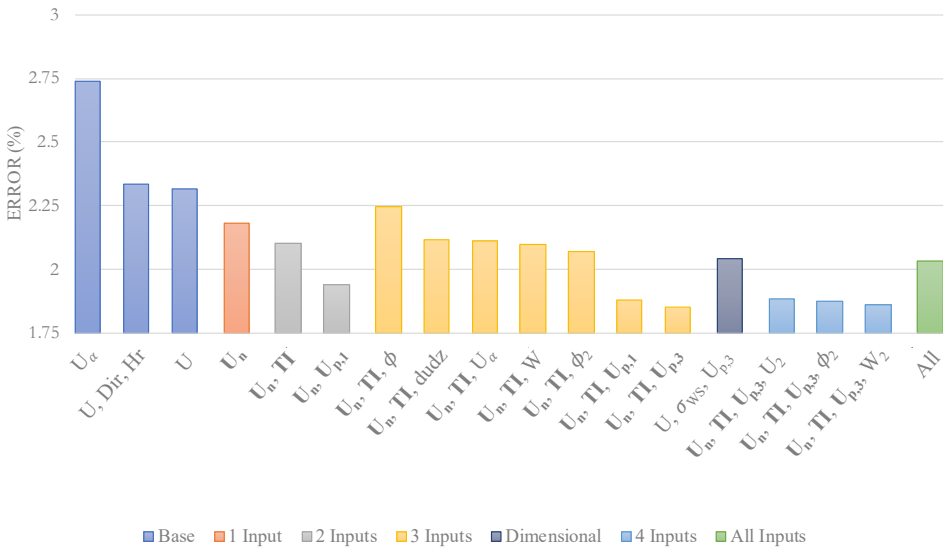


**Figure 3.** MAPE at Perdigão averaged over all heights. Color coding describes the number of input features used for the ANN. Dark blue bar indicates an additional test for comparison. Log law extrapolation (4.3% average MAPE) excluded for clarity.

Two tests were performed to determine whether this improvement in accuracy is derived from feature non-dimensionalization. Because the network performed best at Perdigão with input features of $\boldsymbol{U_n}$, $\boldsymbol{TI}$, and $\boldsymbol{U_{p,3}}$, the same inputs were then given to the network, but in dimensional form (i.e. $U$, $\sigma_U$, and $U_{p,3}$). The dark blue bar in Fig. 3 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 $\boldsymbol{U_n}$, $\boldsymbol{U_{p,3}}$, and $\boldsymbol{TI}$ was analyzed in-depth. In order to determine the exact effects of non-dimensionalization, the same inputs were then given to the network in dimensional form. The results are given in Fig. 4. 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. 4a 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 underpredicting high wind speeds. The non-dimensional network, however, does not have this problem and accurately extrapolates these higher wind speeds.

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

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} \tag{5}$$

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, possibly by forcing it to calculate the amount of shear between reference and extrapolation heights, which is more easily determined with the assistance of $\boldsymbol{TI}$. 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 is seen in Fig. 4c and d, which show the change in uncertainty $\eta$ with height. For a Gaussian variable, $\eta$ can be defined as

$$\eta = 100 \times \frac{\sigma_\varepsilon}{\overline{U}} \tag{6}$$

where $\sigma_\varepsilon$ is the standard deviation of the error (for individual predictions in the testing set) and $\overline{U}$ the mean wind speed. For non-dimensional testing, predicted wind speeds are first transformed back into the dimensional space (i.e. $\boldsymbol{U_n} \rightarrow U$) prior to error calculation in order to find true wind speed extrapolation uncertainty. The total uncertainty, a measure of error variability, is reported in the top row of Fig. 4, but the change in $\eta$ 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. 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 extrapolated wind speeds $> 15ms^{-1}$. The non-dimensional network's uncertainty, meanwhile, continues to decrease as wind speed increases, eventually reaching values as low as 0.5%. This is once again due to the fact that the non-dimensional network is better accounting for the wind shear that is crucial for extrapolation. High wind speeds no longer appear to the network as outliers, allowing the network to better extrapolate much higher wind speeds than otherwise possible. 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. 4e 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 extrapolated wind speeds $< 16ms^{-1}$. Whereas the uncertainty denotes error variability, $MAPE$ denotes overall prediction error. As is clear in Fig. 4a, the dimensional network has an obvious bias at high wind speeds, systematically under-predicting extrapolation wind speed. This is apparent in Fig. 4e, as $MAPE$ increases to more than 10% at higher 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, 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 most sensitive to the choice of input features. This may be due to two factors. First, the site may have flow 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

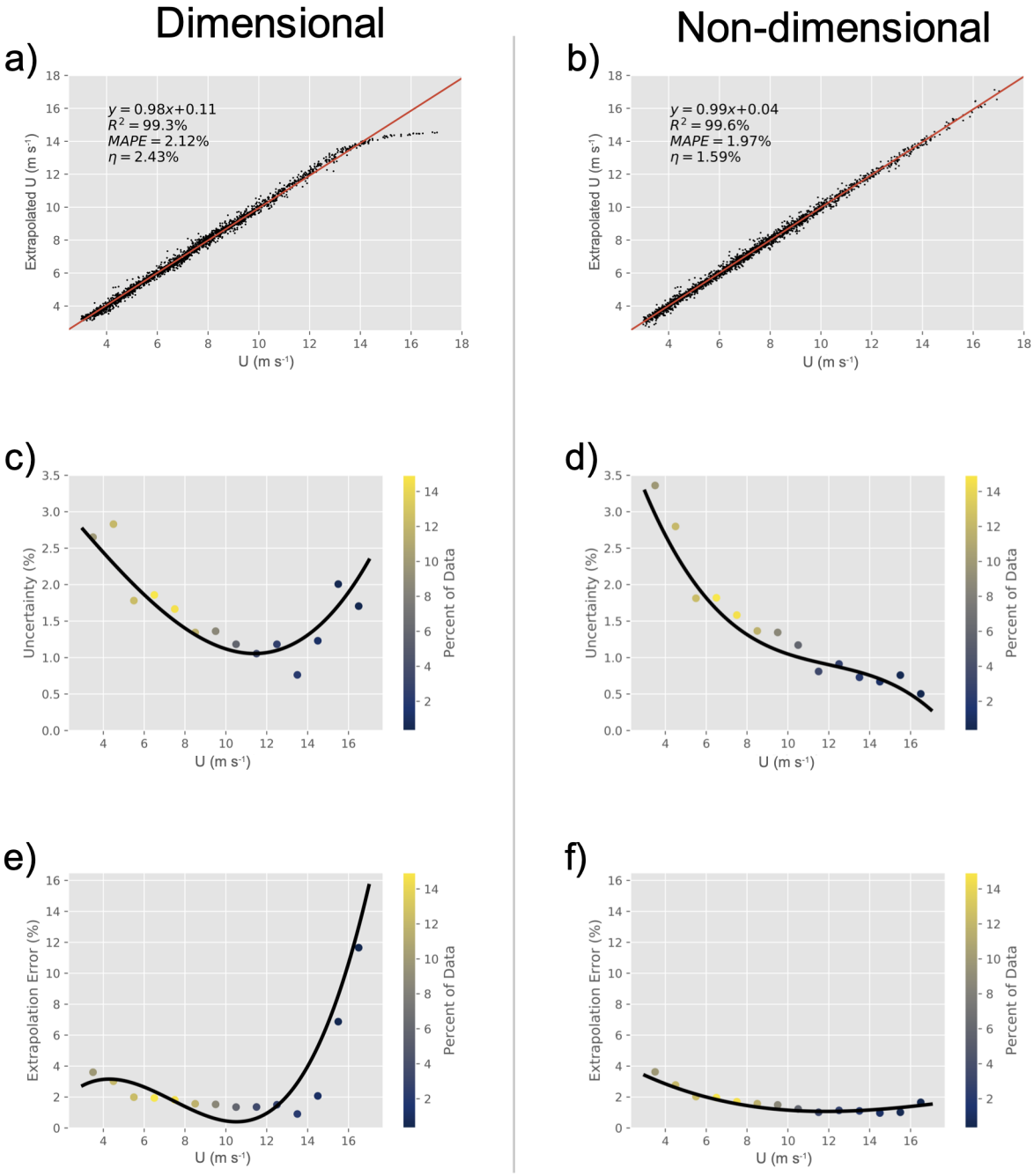


**Figure 4.** Comparison of network performance with dimensional (a, c, e) and non-dimensional (b, d, f) input features for the 160m extrapolation at Perdigão. Top row shows a comparison of true and extrapolated wind speed (at extrapolation height) with the best-fit line, the middle row the change in 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. Colors indicate the percent of data at each binned wind speed for the testing dataset.

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 may be due to the fact that the power law extrapolation performed 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 non-dimensional input features, $\boldsymbol{U_n}$, $\boldsymbol{U_p}$, and $\boldsymbol{TI}$, most reliably help the efficacy of the ANN. Extracting the non-dimensional 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 drastically improves accuracy. This is the only feature that contains information about the flow's history. 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 ($\phi$, $W$, $Dir$) 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 ($dudz$, $U$) provide redundant information. These features typically fail to improve network accuracy, can slow the training process, and are best left out. Lastly, $U_\alpha$ can act as a positive or negative influence on the network because $\alpha$ is dependent on other parameters such as $\boldsymbol{U_n}$, $\boldsymbol{U_p}$, and $\boldsymbol{TI}$. If the power law model is reasonably accurate or has a clear repetitive bias, $U_\alpha$ could be a useful input feature that provides the ANN with a dependable indication of wind shear. Otherwise, it adds misleading noise to the input feature set by thwarting the steering that $\boldsymbol{U_n}$, $\boldsymbol{U_p}$, 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 that the simplification does not remove necessary information. An example is the difference in extrapolation accuracy between $U$ and $\boldsymbol{U_n}$. Before the non-dimensionalization, the ANN has to find a baseline wind speed and predict the vertical wind shear. With non-dimensionalization, the baseline wind speed is a constant and the network is able to exploit possible self-similarity properties of the velocity profile. Whatever information lost during non-dimensionalization 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 have a scarce amount of data (i.e. extrapolated wind speed $> 14ms^{-1}$).

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 forecasting. Recurrent neural networks

should also be utilized to test how alternative 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 that has vast economic implications. It has been shown that standard wind energy vertical extrapolation methods are outdated and can no longer serve their purpose of efficiently predicting and extrapolating meteorological properties accurately under various conditions (Sfyri et al., 2018; Stiperski et al., 2019). This problem can be mitigated by employing machine learning tools that 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 delve into 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, especially on governing dynamical variables, can greatly assist these systems in finding underlying trends that govern atmospheric phenomena.

This study investigated the efficacy of utilizing ANNs for vertical wind speed extrapolation over a variety of terrains and evaluated how various meteorological input feature sets may influence extrapolation accuracy. Various meteorological features were combined to test their effectiveness as ANN inputs. It was found that, on the average, ANN vertical extrapolation error decreases by 15% when using $\boldsymbol{U_n}$ as a singular input feature rather that $U$. Two other extracted non-dimensional features, $\boldsymbol{TI}$ and $\boldsymbol{U_p}$, also led to increased extrapolation accuracy. The accuracy obtained by the ANN was up to 65% and 53% better than that obtained by a log law and power law vertical extrapolations, respectively. Vertical extrapolation error was minimized to as low as 1.06% over 20m, but too many network inputs (many of which are highly correlated) actually caused a reduction in network accuracy. The 160m extrapolation at Perdigão was analyzed in depth to determine the effects of feature non-dimensionalization. In addition to an improved correlation with measured wind speeds, non-dimensionalization led to a decrease in both total extrapolation error and variability, particularly at high wind speeds. The non-dimensional input features created a robust network that improved 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 the current industry standard vertical extrapolation accuracy.

Future studies are planned to investigate feature extraction and selection for wind speed predictions over a variety of timescales using a recurrent neural network. Identification of robust non-dimensional variables is expected to give ANNs 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. Input and target variables are altered for each individual test; example codes used for this study maybe found at https://github.com/dvassall/.

## Appendix A: MAPE Magnitude Invariance

Our goal is to ensure that the loss function's magnitude is invariant regardless of output scaling, allowing a fair comparison between dimensional and non-dimensional networks. For a simple feed-forward neural network with $j$ output nodes, our output error can be defined as $E = \frac{1}{n}\Sigma_c\Sigma_j e_{cj}$ where $c$ is the number of samples in a batch and $e_{cj}$ is the error (given by a user-defined loss function) seen by each output node for each sample in the batch. We use the mean absolute percentage error loss function, meaning that we may define our error metric as

$$e_{cj} = 100\frac{|y_{cj} - \hat{y}_{cj}|}{y_{cj}} \tag{A1}$$

where $y_{cj}$ is the true target output, $\hat{y}_{cj}$ is the predicted target output, and the vertical lines denote the absolute value. For convenience consider a single sample in a single batch (this same analysis can be expanded to multiple samples over multiple batches because of the linear nature of the summation). We will refer to the true and predicted values as $y$ and $\hat{y}$, respectively. We can now define the true and predicted (dimensional) outputs ($y_d$ and $\hat{y}_d$, respectively) as well as the true and predicted non-dimensional outputs ($y_n = {}^{y_d}/_{a}$ and $\hat{y}_n = {}^{\hat{y}_d}/_{a}$, respectively, where $a$ is a non-dimensionalization variable unique to each individual case). We can find the dimensional error $e_d$ to be

$$e_d = 100\frac{|y_d - \hat{y}_d|}{y_d} \tag{A2}$$

Likewise, the non-dimensional error $e_n$ can be written as

$$e_n = 100\frac{|{}^{y_d}/_{a} - {}^{\hat{y}_d}/_{a}|}{{}^{y_d}/_{a}} = 100\frac{|y_d - \hat{y}_d|}{y_d} \tag{A3}$$

proving that the error's magnitude is invariant under non-dimensionalization. This is not true for loss metrics such as mean squared error or mean absolute error, where

$$e_d = (y_d - \hat{y}_d)^2, \quad e_n = \frac{(y_d - \hat{y}_d)^2}{a^2} \tag{A4}$$

and

$$e_d = |y_d - \hat{y}_d|, \quad e_n = \frac{|y_d - \hat{y}_d|}{a} \tag{A5}$$

respectively. By using the $MAPE$ loss function, we are ensuring that the network learns at similar rates when using both dimensional and non-dimensional variables.

## Appendix B: Definitions of Variables

Features are listed in the order that they appear in Section 2.2. Non-dimensional features are in bold. All input variables except $Hr$ are taken from four elevations below extrapolation height.

| | |
|---|---|
| $U_\alpha$ | Extrapolated wind speed based on Eq. 2 ($ms^{-1}$) |
| $U_L$ | Extrapolated wind speed based on Eq. 3 ($ms^{-1}$) |
| $U$ | Streamwise wind speed ($ms^{-1}$) |
| $Dir$ | Wind direction ($-1 \rightarrow 1$) |
| $Hr$ | Hour of the day (cosine curve; $-1 \rightarrow 1$) |
| $\boldsymbol{TI}$ | Turbulence intensity; $\boldsymbol{TI} = {}^{\sigma_U}/_{U}$ ($\sigma_U$ is the standard deviation of streamwise wind speed; both $\sigma_U$ and $U$ taken at a single elevation) |
| $\boldsymbol{U_n}$ | Non-dimensional streamwise wind speed ($U$ at all heights divided by $U$ 20m below the extrapolation height) |
| $\boldsymbol{U_p}$ | Non-dimensional streamwise wind speed from the previous time period (non-dimensionalized the same way as $\boldsymbol{U_n}$) |
| $dudz$ | Vertical wind shear ($dudz = {}^{\partial U}/_{\partial z}$; $s^{-1}$) |
| $\phi$ | Terrain elevation angle from direction of incoming wind speed ($^\circ$) |
| $W$ | 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. The Pacific Northwest National Laboratory is operated for the DOE by Battelle Memorial Institute under Contract DE-AC05-76RLO1830. We express appreciation to Prof. Julie Lundquist and her graduate students at the University of Colorado Boulder for collecting the Wasco lidar data with financial support from the Department of Energy. The CASPER-West data were collected by Mr. Ryan Yamaguchi as a part of the Naval Postgraduate School contribution to CASPER directed by Professor Qing Wang. The Perdigão data are from a lidar deployed by the Institute of Science and Innovation in Mechanical and Industrial Engineering (INEGI), the operation of which was overseen by Dr. Jose Carlos Matos.

## 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.

Akish, E., Bianco, L., Djalalova, I. V., Wilczak, J. M., Olson, J. B., Freedman, J., Finley, C., and Cline, J.: Measuring the impact of additional instrumentation on the skill of numerical weather prediction models at forecasting wind ramp events during the first Wind Forecast Improvement Project (WFIP), Wind Energy, 2019.

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.

Bianco, L., Djalalova, I. V., Wilczak, J. M., Olson, J. B., Kenyon, J. S., Choukulkar, A., Berg, L. K., Fernando, H. J., Grimit, E. P., Krishnamurthy, R., et al.: Impact of model improvements on 80 m wind speeds during the second Wind Forecast Improvement Project (WFIP2), Geoscientific Model Development (Online), 12, 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.

Bodini, N., Lundquist, J. K., Krishnamurthy, R., Pekour, M., Berg, L. K., and Choukulkar, A.: Spatial and temporal variability of turbulence dissipation rate in complex terrain, Atmospheric Chemistry and Physics, 19, 4367–4382, 2019.

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.

Dupré, A., Drobinski, P., Alonzo, B., Badosa, J., Briard, C., and Plougonven, R.: Sub-hourly forecasting of wind speed and wind energy, Renewable Energy, 145, 2373–2379, 2020.

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.

Holmes, J. D.: Wind loading of structures, CRC press, 2018.

Hsieh, W. W. and Tang, B.: Applying neural network models to prediction and data analysis in meteorology and oceanography, Bulletin of the American Meteorological Society, 79, 1855–1870, 1998.

Islam, M. S., Mohandes, M., and Rehman, S.: Vertical extrapolation of wind speed using artificial neural network hybrid system, Neural Computing and Applications, 28, 2351–2361, 2017.

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.

Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

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.

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.: Lidar-CU WindCube V1 Profiler, Wasco Airport-Reviewed Data, Tech. rep., Atmosphere to Electrons (A2e) Data Archive and Portal, Pacific Northwest National Laboratory; PNNL, 2017.

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.

Markou, H. and Larsen, T. J.: Control strategies for operation of pitch regulated turbines above cut-out wind speeds, in: 2009 European Wind Energy Conference and Exhibition, EWEC, 2009.

McGovern, A., Lagerquist, R., John Gagne, D., Jergensen, G. E., Elmore, K. L., Homeyer, C. R., and Smith, T.: Making the black box more transparent: Understanding the physical implications of machine learning, Bulletin of the American Meteorological Society, 100, 2175–2199, 2019.

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.

Olson, J. B., Kenyon, J. S., Djalalova, I., Bianco, L., Turner, D. D., Pichugina, Y., Choukulkar, A., Toy, M. D., Brown, J. M., Angevine, W. M., et al.: Improving wind energy forecasting through numerical weather prediction model development, Bulletin of the American Meteorological Society, 100, 2201–2220, 2019.

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.

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.

Türkan, Y. S., Aydoğmuş, H. Y., and Erdal, H.: The prediction of the wind speed at different heights by machine learning methods, An International Journal of Optimization and Control: Theories & Applications (IJOCTA), 6, 179–187, 2016.

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.

Wharton, S. and Lundquist, J. K.: Atmospheric stability affects wind turbine power collection, Environmental Research Letters, 7, 014 005, 2012.

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.