Response to Reviewer 1

Dear Reviewer,

Thank you for your thorough review of the manuscript. In response to your comments, we have substantially revised the manuscript to make it more clear and focused. The majority of the changes have been made to the results section, where we now focus on showing improvements in lidar TI, rather than improvements in power prediction. We have also reduced the number of figures and aimed to make figure and table captions more descriptive. Another major change is the removal of the BAO dataset from the manuscript, as we did not feel confident in the quality of the lidar data from the site. A point-by-point response to your comments is given below.

The manuscript states that it presents a new method of estimating turbulence for lidar and as such represents a new contribution. Overall, it is difficult to determine how important this technique is because as far as it has been possible to determine the corrected turbulence is not evaluated/presented except after being filtered through power prediction. These improvements do seem to be very small – can you comment on whether this is significant? And if so which part of the model streams are necessary?

Response: Results from the TI correction are now shown in Section 5 of the revised manuscript. Several metrics are given to indicate improvement in the lidar TI, including MAE, regression line slope, and changes in TI error sensitivity to different external conditions. It was determined that different parts of the model stream are needed depending on the stability conditions, which is discussed on p. 11, Lines 16-21.

Much of the introductory material is qualitative and could be much improved and likely made briefer by using quantitative metrics. So for example from the abstract it is not possible to evaluate how good the new method is or how much effort is needed to implement it.

Response: We have now included a quantitative indication of the improvement of lidar TI with L-TERRA in the abstract and have also given this information in Tables 3 and 4 in the text.

This conversational style continues through the Introduction which could be much reduced or improved. For example, 'turbulence' is not defined – so when comparing turbulence from cup anemometers or sonic anemometers versus lidar what is really being compared?

Response: Turbulence intensity has now been defined in the introduction section. We also acknowledge the difference in how TI is calculated for cup vs. sonic anemometers in Section 3.1 (p. 8, Lines 7-11).

In the Background section there needs to be some reorganization. If there is going to be a general introduction to lidar technology this is where different lidar systems should be compared – with a focus on how 'turbulence' is derived by different systems.

Response: The background section has been revised. Information on different lidar systems and methods for deriving turbulence is now included in Section 2.1 – Lidar Technology.

The section on errors in lidar data is probably too broad and needs to be quantitative. Which of these errors will dominate wind speed and turbulence estimation? If not maybe it could simply be a table with references. The WC scanning circle (p4, l5-10) is surely dependent on the instrument and not a general quantity.

Response: As our goal was to inform the reader qualitatively about the different sources of error, we have decided to keep this section intact. We agree that it's important to determine which sources of error will dominate wind speed and turbulence estimation, but this is a complex question, as it depends on atmospheric conditions, site conditions, and the lidar technology and scanning strategy used.

Similarly, the section on correcting lidar turbulence gives a qualitative overview. If this cannot be quantitative, which would be best approach, then a table giving the necessary inputs, output and advantages and disadvantages would be helpful.

Response: We have decided to keep this information in the text, as the manuscript already contains five tables and much of the information on how the techniques are implemented cannot be fit into a table. We appreciate the suggestion.

In section 7 it is not clear why the step with the power law is needed – this surely introduces much larger errors than can be corrected for later?

Response: The power law is used to determine the shear parameter α, which is then used as a proxy for atmospheric stability and as an input variable for the machine learning models discussed in Section 5.2. Although there are certainly limitations to using the power law, our goal is merely to relate what the lidar is measuring to the errors in the TI.

Is it correct then that you are defining turbulence as the standard deviation of u/u for lidar, cups and sonics? (p7, l26). If so probably best to state this upfront. You say the process is similar for all three but it can't be – no coordinate rotation for cups?

Response: It has now been stated in Section 3.1 (p. 8, Lines 7-11) that the way that TI is calculated for cup anemometers is different than how it is calculated for sonic and lidar measurements.

It is my understanding of what you have written that you apply corrections for 1) instrument noise in the form of a spike filter 2) volume averaging and 3) variance contamination and then use machine learning to train (TI error?) on predictor variables including shear parameter, mean wind speed etc.

Response: This is correct and is explained in Figure 2 of the manuscript.

The datasets are 6 months from ARM site cf 60 m tower data, 2 months from BAO vs 300 m tower and 7 months from a wind farm.

Response: The BAO dataset has been removed from the manuscript, so we now use approximately 6 months of data from both the ARM site and the wind farm.

You had sonic data and used the shear parameter to classify stability, why? (Table 2)?

Response: We agree that using sonic anemometer data to calculate the Monin-Obukhov length or Richardson number would be a more ideal way to classify stability. However, we wanted to use information from a stand-alone lidar to classify stability, which is why we decided to use the shear parameter as a proxy for stability (p. 10, Lines 23-25).

Table 3/4 is unclear what it is showing and what has been done? What is the MARS model? is this different from Terra?

Response: The information in Table 3 has now been split into three tables in the revised manuscript: Table 5 contains optimal model combinations for L-TERRA, and Tables 3 and 4 contain TI error metrics for the ARM site and wind farm, respectively. In addition, the physics-based corrections and machine-learning corrections are now discussed in two separate sections (Sections 5.1 and 5.2).

Please work on these Table and Figure captions. It is really difficult to understand what they pertain to.

Response: The table and figure captions have been revised.

Figure 1: Not sure why this is here or why it is needed.

Response: Figure 1b (power sensitivity to TI) has now been removed from the manuscript. Figure 1a (TI-dependent power curves) has been retained to indicate the importance of measuring TI to determine turbine power production.

Figure 3/4. Quality needs to be improved.

Response: The quality of Figure 3 has been improved. Figure 4 has been removed from the manuscript, as the BAO data are now longer discussed.
Figure 5. Not sure why these histograms are needed. Which instruments are these from?

Response: This figure has been removed from the manuscript.

Figure 6. Not sure why these graphs are needed. What is the corrected turbulence here?

Response: This figure has been removed from the manuscript.

Figure 7. Not really sure why the regression lines are needed. Are these the TI corrected data?

Response: This figure has been removed from the manuscript. The raw and corrected TI for the ARM site are now shown in Figure 4. To make the figures more clear, the number of stability classes has been reduced to three, and different plots are used to show TI data for different stability classes.

I was surprised when I got to the end of the Tables and Figures that no results are presented for the turbulence-estimates given that is the main topic of the paper?

Response: We agree that results for the turbulence estimates should be shown in the paper. We now focus solely on improvement of the lidar TI and plan to discuss applications of the TI correction (e.g., on power prediction) in future work.

Section 5 is descriptive of the general behavior of turbulence at the sites – is it needed?\ There did not seem to be anything here that was unexpected so it was not clear why it is present. When I look at Figure 7 I think you are showing that there is better agreement between the sonic TI and the (uncorrected?) lidar TI when the wind shear is lower?

Response: Section 5 has now been removed from the manuscript.

The results section is very unclear. A better approach would be to indicate how the results were obtained rather than stating 'optimal model combinations are shown in Table 3' – this table refers to MAE in kW before and after (training with Terra?) and indicates a reduction in MAE – so not turbulence? Where are the turbulence results? It makes me wonder if I am missing the point of the paper. How important is this? If the overall power is 2MW – then is a reduction in MAE from 2.16 to 1.77 kW important? Actually now I read this I realise I don't follow this at all. Is this like a reduction in error from 0.14% to 0.11% - is it worth this effort? Where does the error in reduction come from? Is this on the average power over one year or ? Unfortunately I was unable to determine the results that lead to the statement 'L-TERRA improves TI estimates' – how would I see that?

Response: We now focus only on the results of the turbulence correction and do not translate the improvements in TI to improvements in power prediction. We agree that this translation to power prediction made the results section unclear. In addition, we now show changes in TI error sensitivity that result from application of L-TERRA (Figure 5), as the reduction of TI error sensitivity to external parameters is another way to demonstrate the benefits of L-TERRA.

Overall the paper had an interesting premise but it needs a major overall to clarify and show the results pertaining to turbulence and to remove unnecessary material. The error reduction approach has to be systematically compared to other methods and to show the effort level required to obtain it to be of utility. As it stands it would not be possible for anyone to reproduce this work.

Response: We have significantly revised the manuscript to clarify the techniques used and focus solely on the correction of lidar turbulence. It would be difficult to compare the results of L-TERRA quantitatively, as most of the current turbulence error reduction techniques discussed in the manuscript use scanning lidars or data that is not readily available from vertically profiling lidars. However, we greatly appreciate your suggestions to make the manuscript more clear and concise.

Response to Reviewer 2

Dear Reviewer,

Thank you for your thorough and extremely helpful review of the manuscript. In response to your comments, we have substantially revised the manuscript and focused on improving the lidar TI, rather than reducing errors in power prediction. A more complete description of the physics-based corrections and machine learning techniques is now included, as well as a thorough and clear discussion of the results from both the physics-based and machine learning corrections in L-TERRA. A point-by-point response to your comments is given below.

The following questions should be answered in the paper:

1. How much of the correction is done by the physical modules and how much by the machine learning module?

Response: The application of the physics-based corrections and the machine learning techniques are now separated into two different sections (Sections 5.1 and 5.2), so the effects of each module on the resulting TI estimates can now be seen clearly.

2. What is the remaining error, after correction by the physical modules, due to? Is there any systematic pattern in the remaining error? If not (in extreme case, if the remaining error looks random) can a machine learning approach really make an improvement?

Response: A sensitivity analysis was conducted in Section 5.2, and the use of the sensitivity analysis results to determine patterns in the remaining error is discussed on p. 12, Lines 24-30, and p. 13, Lines 32-25–p. 14, Lines 1-4.

3. To which extent the machine learning module can improve the correction from the physical correction modules? In other words, can the improvement compared to previous work cited in section 3 be quantified and clearly demonstrated?

Response: The physical corrections and machine learning methods are now discussed separately in the revised manuscript. It was determined that training a random forest or MARS model on one dataset and testing on a different dataset led to an increase in MAE and decrease in R^2 values. While training and testing on the same site did decrease MAE values, it still increased R^2 values.

It would indeed be useful to compare our approach to the results from previous methods, but it would be difficult to make this comparison in reality. Most of the

previous methods discussed require information that is not readily available from a vertically profiling lidar.

4. In the introduction, the L-TERRA algorithm is presented as a method easier to apply than other methods for lidar turbulence measurement correction proposed so far (p2, 25). However, in the end, the method does not look really much easier (and maybe even more complex) to implement since it re-uses or adapts the methods qualified as complex to the WC, combine them and add the machine learning module. The simplicity of the approach needs to be further demonstrated.

Response: Section 3.6 (pp. 9-10) has been added to compare the evaluation of L-TERRA to other methods described in the text.

Furthermore, a more straight forward evaluation of the results, like a direct comparison of the lidar measured turbulence intensity to the sonic measurements before and after correction (as done in figure 11) would lead to clearer conclusions. The conversion of the TI/TI error into power/power error is very interesting to demonstrate the impact of the TI measurement error on power prediction (as done in section 5), but it tends to confuse the analysis objectives when it is used in the results (like it is now done in section 6.2.).

Response: In the revised manuscript, direct comparisons of the TI have been made both before and after L-TERRA has been applied (Section 5). To make the paper more clear, we have decided to reserve the implications of L-TERRA on power prediction for future work.

Detailed comments:
(p: page; l: line)

P2; l.27-32: this paragraph is misleading and/or misplaced. It gives the conclusions of the analysis. I suggest removing it.

Response: This paragraph has been removed.

P7, l5: is it 4s second scan or 5 sec scan?

Response: The actual accumulation time of the WINDCUBE lidar used in this work was just under 1 second, so a full scan actually took closer to 4 seconds. This has been clarified in the revised manuscript.

P8, l7: "Two methods were evaluated. . ." The flow chart in fig. 2 shows 4 methods to take care of the effect of volume averaging (2 for each type of wind speed). Does that mean only 2 of them have been tested? Which ones: "spectral filtering 1" and "spectral filtering 2"?

Response: It has now been clarified that only the model path that incorporates the *u, v*, and *w* components was tested, as not all vertically profiling lidars include the radial velocity measurements in their output files (p. 7, Lines 7-10). The difference between Spectral Fit 1 and Spectral Fit 2 has also now been clarified in the manuscript (p. 9, Lines 1-4).

P8, l11-12: could you please provide explanations on how you have applied the method from Krishnamurthy to DBS scans?

Response: A description of applying the structure function technique to DBS data is now described in Section 3.3 (p. 8, Lines 22-26).

P8, l17-19: could you please provide explanations on how you have applied those techniques to a 5 beam configuration?

Response: An explanation for applying the six-beam technique to a five-beam configuration has been added to Section 3.4 (p. 9, Lines 8-12).

P8, l24: "there is still some error . . ." Why? What is the remaining error due to? Does it mean the physical correction models applied previously are not good enough? The assumptions are not verified?

Response: This section has been revised so that it now simply describes the content of the machine learning module in L-TERRA (Section 3.5).

P9, l1: for the paper to be comprehensive by itself, it would be good to have a short description of each of the 3 machine learning methods.

Response: Random forests and the MARS model are now described in more detail in the revised manuscript (p. 9, Lines 15-22). References to support vector regression were removed in the revised manuscript, as this model performed poorly on the test datasets in comparison to MARS and random forests.

P9, l21-22:
1. Some of these results are rather surprising to me: internal temperature and pitch. Could you please comment on those and provide some information regarding the range of each variable and its correlation with the TI?

Response: We agree that the high sensitivity of TI error to internal temperature and pitch was a bit surprising. After applying the new version of L-TERRA in the revised version of the paper, this high sensitivity to internal temperature and pitch was no longer evident.

2. TI and sigma_w were not correlated?

Response: TI and sigma_w were correlated, though not extremely strongly (correlation coefficient = 0.363). We used a correlation coefficient of 0.5 to discriminate between weakly and strongly correlated variables, so both TI and sigma_w were retained in the list of predictors.

3. Have you performed the same sensitivity analysis for the other flat site to see if you get the same final predictor variables?

Response: We did perform the same sensitivity analysis for the other flat site and determined that while some variables had similar sensitivities at both sites, others had significantly different sensitivities. This is discussed on p. 13, Lines 6-10 in the revised manuscript.

P10, section 4

1. In the introduction you mention 2 flat sites and one semi-complex.
Which one of the three sites described here is the semi-complex one? My guess is BAO, but needs to be clearly sated as it can influence the results of the lidar measurements.

Response: The BAO was indeed the semi-complex site. References to the BAO have been removed from the revised manuscript, as we were not confident in the lidar data quality from the site.

2. In the semi complex site, were the lidar measurements corrected for the effect of complex terrain (e.g. FCR) or the same wind speed reconstruction algorithm was used at all sites?

Response: The lidar measurements were not corrected for the effect of complex terrain.

P12, section 5 The main reason for the poor results at BAO is probably the effect of the terrain on the flow. In complex terrain, the assumption of horizontal homogeneity is not verified, then the reconstructed mean wind speed includes some error and therefore does not correlate with the sonic measurements as well as lidar measurements in flat terrain. If the mean wind speed comparison to the sonic is poor, the TI comparison is also expected to be poor.

Response: We agree that the effect of complex terrain on the flow, in addition to the low aerosol count at the BAO, adversely affected the accuracy of the lidar measurements at the site. Thus, we have removed the BAO data from our analysis.

P13, l19-20: Does this mean that the correction algorithm for the radial wind speed (right part of the flow chart In figure 2) has not been tested yet (or at least the results are not included in this paper)? This sounds in contradiction with p13, l22-23 (and p6, l29-30) stating that "all possible combinations . . . were evaluated". Could you please clarify which of the process presented in flow chart in fig 2 have actually been used for the results presented and discussed in this paper?

Response: The correction algorithm for radial wind speed has not yet been tested extensively and was not included in this work. This is now stated on p. 7, Lines 7-10 of the revised manuscript.

P13, l 30-31: Only the optimal combinations for each site are presented in table 3.

1. How much difference was there between the different combinations? Was it significant?

Response: Several model combinations produced similar MAE values for each site. This is now stated on p. 11, Lines 9-11.

2. Is one of the combinations more robust than the others? i.e. could you , based on this analysis, recommend one combination or is the idea to always try all of them and pick the smallest error? (This is maybe to be included in the discussion in section 6).

Response: As minimizing the MAE produced several similar "optimal" model combinations, we suggest that looking at other parameters might be more useful for determining the ideal model combination (p. 11, Lines 11-14). However, for this initial evaluation of L-TERRA, minimizing the MAE is helpful for finding some baseline model combinations for comparison with the raw TI.

p.14, l13-28:

1. From this analysis, it sounds like the cross contamination effect is the most important. How much the correction for this effect change the lidar does measured TI? And how much does the machine learning correction part change the lidar measured TI?

Response: We still believe the cross-contamination effect is important, although we have now phrased the effect in terms of changes in the scales of vertical velocity across the lidar scanning circle (p. 13, Lines 32-35-p. 14, Lines 1-4). For the physics-based corrections, the variance contamination module reduced TI estimates by an average of 0.15% for stable conditions and 0.8% for unstable conditions. As discussed in Section 6.2, machine learning techniques increased MAE and scatter in the TI estimates and likely did not have a large effect on reducing variance contamination. We are currently examining ways to improve the physics-based variance contamination module to further improve TI estimates.

2. Figure 11 shows that the slopes in the linear regression are usually getting closer to 1 after application of the L-TERRA correction, which mean the mean error is reduced. But the scatter is increased (R^2 is lower), so the improvement is actually mitigated. Moreover, this shows that the method does not necessarily provide better estimate of every 10 minute value of TI measure by the lidar, whereas it was demonstrated in section 5 that is what was needed.

Response: This increase in scatter was mitigated through application of the stability-dependent version of L-TERRA (see Tables 3 and 4 in the revised manuscript). However, we agree that the R^2 values still indicate a large amount of scatter in the data and suggest that this scatter occurs because we are not completely capturing all the physics that affect TI error (p. 12, Lines 4-7).

P15, section 6.2: only MARS and RF are discussed. Support Vector Regression was not tested?

Response: Support vector regression was initially tested but produced poor results in comparison to random forests and the MARS model, so we have not included a discussion of support vector regression in the manuscript. We have also removed references to support vector regression in the section on machine learning (Section 3.5).

An Error Reduction Algorithm to Improve Lidar Turbulence Estimates for Wind Energy

Jennifer F. Newman[1] and Andrew Clifton[2]

[1]National Wind Technology Center, National Renewable Energy Laboratory, Golden, CO, 80401, USA
[2]Power Systems Engineering Center, National Renewable Energy Laboratory, Golden, CO, 80401, USA

Correspondence to: Jennifer F. Newman (Jennifer.Newman@nrel.gov)

Abstract.

Remote sensing devices such as lidars are currently being investigated as alternatives to cup anemometers on meteorological towers. Although lidars can measure mean wind speeds at heights spanning an entire turbine rotor disk and can be easily moved from one location to another, they measure different values of turbulence than an instrument on a tower. Current methods for improving lidar turbulence estimates include the use of analytical turbulence models and expensive scanning lidars. While these methods provide accurate results in a research setting, they cannot be easily applied to smaller, ~~commercially available~~ vertically profiling lidars in locations where high-resolution sonic anemometer data are not available. Thus, there is clearly a need for a turbulence error reduction model that is simpler and more easily applicable to lidars that are used in the wind energy industry.

In this work, a new turbulence error reduction algorithm for lidars is described. The algorithm, L-TERRA, can be applied using only data from a stand-alone ~~commercially available~~ vertically profiling lidar and requires minimal training with meteorological tower data. The basis of L-TERRA is a series of physics-based corrections that are applied to the lidar data to mitigate errors from instrument noise, volume averaging, and variance contamination. These corrections are applied in conjunction with a trained machine-learning model to improve turbulence estimates from a vertically profiling WINDCUBE v2 lidar. The lessons learned from creating the L-TERRA model for a WINDCUBE v2 lidar can also be applied to other lidar devices.

L-TERRA was tested on data from ~~three sites - two in flat terrain and one in semicomplex terrain.~~ two sites in the Southern Plains region of the United States. The physics-based corrections in L-TERRA ~~significantly reduced errors in lidar turbulence at all three sites , even when the~~ reduced the mean absolute error by nearly 20% at both sites and significantly reduced the sensitivity of lidar turbulence errors to atmospheric stability. The accuracy of machine-learning ~~portion of the model was trained on one site and applied to a different site. Errors in turbulence were then related to errors in power through the use of a power prediction model for a simulated 1.5MW turbine. L-TERRA also reduced errors in power significantly at all three sites, although moderate power errors remained for periods when the mean wind speed was close to the rated wind speed of the turbine and periods when variance contamination had a large effect on the lidar turbulence~~ methods in L-TERRA was highly dependent on the input variables and training dataset used, suggesting that machine learning may not be the best technique for

reducing lidar TI error. Future work will include the use of a lidar simulator to better understand how different factors affect lidar turbulence error and to determine how these errors can be reduced using information from a stand-alone lidar.

1 Introduction

As turbine hub heights increase and wind energy expands to complex and offshore sites, new measurements of the wind resource are needed to inform decisions about site suitability and turbine selection. Currently, most of these measurements are collected by cup anemometers on meteorological (met) towers. Met towers are fixed in location and typically only collect measurements up to and including the height corresponding to the turbine hub height. However, the measurement of wind speeds across the entire turbine rotor disk is extremely important for power estimation (e.g., Wagner et al., 2009), particularly as modern turbines increase in size. In addition, met towers are expensive to construct and maintain; the estimated cost for installing and maintaining an 80m land-based met tower for a 2-year campaign is €92,000 (≈ 105,000 USD; Boquet et al., 2010). In response to the limitations of met towers for wind energy, remote sensing devices such as lidars (light detection and ranging) have been proposed as potential alternatives to cup anemometers on towers. Lidars are now frequently used in the research community (e.g., Barthelmie et al., 2013; Stawiarski et al., 2013; Fuertes et al., 2014; Sathe et al., 2015b), and acceptance of lidars in the wind energy community is increasing. ~~In fact, the~~ The use of remote sensing devices for power performance testing in flat terrain is discussed in Annex L of the most recent draft version of IEC 61400-12-1 (International Electrotechnical Commission, 2013).

While lidars are capable of measuring mean wind speeds at several different measurement heights (e.g., Sjöholm et al., 2008; Peña et al., 2009; Barthelmie et al., 2013; Sathe et al., 2015b), they measure different values of turbulence than a cup or sonic anemometer (e.g., Sathe et al., 2011; Newman et al., 2016b). Turbulence, a measure of small-scale fluctuations in the atmospheric flow, is an extremely important parameter in the wind energy industry. Turbulence measurements are used to classify potential wind farm sites and select suitable turbines (International Electrotechnical Commission, 2005) and can also impact power production. Figure 1 shows the response of a modeled 1.5MW turbine to different 10min mean hub-height wind speeds and levels of hub-height turbulence intensity (TI)~~for the same mean hub-height wind speed~~, defined as the standard deviation of the streamwise wind speed divided by the average wind speed over a 10min period and multiplied by 100%. The power produced by the turbine is profoundly impacted by the level of turbulence, particularly near the rated wind speed of the turbine. Because of the paramount importance of turbulence measurements to the wind energy industry, lidars must be able to accurately measure turbulence to be considered a viable alternative to met towers. The inability of lidars to accurately measure turbulence is currently one of the main barriers to replacing met towers with lidars.

In this work, a new turbulence error reduction model, the Lidar Turbulence Error Reduction Algorithm (L-TERRA), was developed for the WINDCUBE v2 (WC) vertically profiling lidar. The model combines ~~physical~~ physics-based corrections, such as a spectral correction, with machine-learning techniques to improve lidar turbulence estimates. ~~These estimates are then related to changes in power prediction through the use of a power prediction model~~While the physics-based corrections can be applied using data from the lidar itself, the machine-learning portion of L-TERRA requires training with a collocated lidar/met

tower dataset. Unlike other methods for improving lidar turbulence estimates, L-TERRA is a simple method that can be easily applied to ~~commercially available lidars. In this paper, the development and initial testing~~ vertically profiling lidars. The goal of L-TERRA ~~are discussed.~~

~~Results indicate that L-TERRA significantly improves lidar TI and power prediction estimates at three different sites, even when the model is trained at one site and applied to data at a different site. Large errors in power prediction still remain for wind speeds near the rated wind speed of the turbine, where the power curve is extremely sensitive to errors in TI , and for low-shear, high-TI conditions, where variance contamination has a strong impact on lidar TI error. Future work will include the use of a lidar simulator to refine the corrections in L-TERRA and~~ is to bring lidar TI estimates closer to the ~~expansion of L-TERRA to different lidar models and configurations~~values of TI that would be measured by a cup anemometer on a tower. Although cup anemometers are affected by overspeeding (e.g., Kaimal and Finnigan, 1994) and mast distortion (e.g., Wyngaard, 1981), they provide sufficient information for wind resource assessment and power performance testing and are the current instrument of reference for wind energy.

The manuscript is organized as follows. Section 2 outlines the main factors that affect lidar turbulence estimates and current methods for improving turbulence estimates. ~~Basic descriptions of~~ A basic description of the different modules in L-TERRA ~~and the power prediction model used in this work are~~ is given in Section 3. The data sets used to train and test L-TERRA are discussed in Section 4~~and the atmospheric conditions present at the different sites are compared. The effects of lidar turbulence error on power prediction at the different sites are described in Section ??. Section ?? includes performance metrics of~~ . Results from L-TERRA are discussed in Section 5, and a sensitivity analysis is conducted to determine the effects of site conditions on lidar TI error both before and after L-TERRA ~~for all three sites, and atmospheric conditions under which the model requires improvement are highlighted~~has been applied. Conclusions and plans for future work are discussed in Section 6.

2 Background

Although lidars are frequently used in wind energy studies (e.g., Peña et al., 2009; Krishnamurthy et al., 2013; Clifton et al., 2015; Wharton et al., 2015; Newsom et al., 2015), they typically measure different values of turbulence than a cup or sonic anemometer (e.g., Sathe et al., 2013; Newman et al., 2016b). In this section, the factors that cause these turbulence discrepancies are discussed. In addition, current methods for reducing turbulence measurement ~~error~~ errors from lidars are highlighted. Throughout this work, the process of "correcting" lidar turbulence refers to techniques that are used to bring lidar turbulence estimates closer to the turbulence that would be measured by a cup or sonic anemometer and "error" is used as a synonym for "difference".

2.1 Lidar technology

Lidars emit laser light into the atmosphere and measure the Doppler shift of the backscattered energy to estimate the mean wind velocity of volumes of air. Laser light from Doppler lidars is typically scattered by aerosol particles in the atmosphere, which are normally prevalent in the boundary layer (Emeis, 2010). For pulsed Doppler lidars, the time series of the returned

signal is split into blocks that correspond to range gates and processed to estimate the average radial wind speed at each range gate (Huffaker and Hardesty, 1996). In contrast, continuous wave lidars focus the laser beam at different distances from the lidar to estimate wind speeds at different ranges (Slinger and Harris, 2012). ~~The sign and magnitude of the radial wind speed are determined from the Doppler shift of the returned signal with respect to the original signal~~

Vertically profiling lidars, which are commonly used in wind energy, involve scanning a cone directly above the lidar to derive the u, v, and w velocity components. If the atmosphere is horizontally homogeneous in the area enclosed by the cone, the radial velocity measured by the lidar, v_r, can be related to the three-dimensional wind components as follows (Weitkamp, 2005):

$$v_r = u \sin\theta \cos\phi + v \cos\theta \cos\phi + w \sin\phi, \tag{1}$$

where θ is the azimuthal angle of the lidar beam, measured clockwise from north, and ϕ is the elevation angle of the lidar beam, measured from the ground. Typically, a raw time series of u, v, and w is derived from Eq. 1 and these raw wind speed components are used to calculate turbulence parameters. A different method involves taking the variance of Eq. 1 and combining the radial velocity variance values from multiple beam positions to directly estimate the u, v, and w variance components (Sathe et al., 2015b).

One lidar that is frequently used in the wind energy industry is the WC, manufactured by Leosphere (Orsay, France). The WC employs a Doppler-Beam swinging (DBS) (e.g., Strauch et al., 1984) technique to estimate the three-dimensional wind vector wherein an optical switch is used to point the laser beam toward the four cardinal directions (north, east, south, and west) at an angle of 28° from zenith ~~. The WC~~ and Eq. 1 is used to derive a time series for u, v, and w. The WC used in this work also includes a vertical beam position for a direct measurement of the vertical velocity. The WC accumulates measurements at each beam position for just under one second, such that a full scan takes approximately 4–5 seconds. However, velocity data from the WC are updated each time new information is obtained (i.e., every time the beam moves to a different position), leading to an output frequency of 1 Hz.

2.2 Errors in lidar data

In Doppler wind lidars, instrument noise results from factors such as the limited amount of aerosol scatterers in the probe volume (Lenschow et al., 2000) and spontaneous radiation emissions from the laser (Chang, 2005). Instrument noise increases the variability of the radial wind speeds measured by the lidar, which artificially increases the turbulence estimates. In contrast, volume averaging decreases the turbulence estimated from the lidar. To obtain a reasonable estimate of the radial velocity, lidars require backscatter data from a large number of scatterers within a probe volume. For the WC, the probe volume measures 20 m along the beam and is negligibly small in the cross-beam and vertical directions. The probe volume acts as a low-pass filter, effectively filtering out all turbulent motions that occur on spatial scales smaller than 20 m. The probe volume is a trade-off between spatial resolution and data accuracy; if the probe volume were smaller than 20 m, fewer data points would be available to estimate the radial velocity, and there would be ~~more~~ a higher amount of uncertainty in the measurements.

The ~~WC, like most other commercial lidars, collects measurements around a scanning circle to estimate the three-dimensional wind vector. At each beam position, the WC obtains an estimate of the line-of-sight velocity, which, for the off-vertical beam positions, contains contributions from all three wind components. If it is assumed~~ scanning strategy used by a lidar can also induce errors in the turbulence estimates. For example, the DBS technique used by the WC requires the assumption that the instantaneous flow field is uniform across the scanning circle~~, then the line-of-sight velocities can be combined to estimate the u, v, and w wind components~~. However, this assumption is generally not true in turbulent flow, when the wind field changes significantly in both space and time (e.g., Wainwright et al., 2014; Lundquist et al., 2015). As the WC scanning circle has a diameter of 106 m at a measurement height of 100 m above ground level (AGL), it is likely that the instantaneous flow field changes in space, even in flat terrain. This changing flow field across the lidar scanning circle introduces additional terms into the variance calculations in a phenomenon known as variance contamination (e.g., Sathe et al., 2011; Newman et al., 2016b). This effect contaminates the true value of the velocity variance and can cause the lidar to measure higher values of turbulence than a cup or sonic anemometer.

2.3 Current methods for correcting lidar turbulence

Several data processing techniques and state-of-the art measurement configurations have already been developed for acquiring turbulence measurements from lidars (Sathe et al., 2015a). However, many of these measurement configurations require expensive scanning lidars or the fitting of turbulence models that are technically only valid under neutral atmospheric conditions. These techniques are applicable in a research setting, but largely require more instrumentation and measurement data than are typically available during a wind resource assessment.

2.3.1 Fitting a turbulence model

One method for correcting lidar turbulence includes modeling the spatial averaging effects of the lidar probe volume. This method involves convolving the true radial velocity field with a spatial weighting function that is controlled by the lidar beam pattern (e.g., Sjöholm et al., 2009; Sathe et al., 2011). Spatial weighting functions for both pulsed and continuous wave lidars are relatively straightforward (e.g., Sonnenschein and Horrigan, 1971). However, modeling the true velocity field requires knowledge of the three-dimensional turbulence structure, which can be described by the spectral velocity tensor, Φ_{ij}.

The spectral velocity tensor can be modeled through use of the Mann (1994) turbulence model, as in Sjöholm et al. (2009), Mann et al. (2010), Sathe et al. (2011), and others. Fitting the model requires three parameters: a turbulence dissipation rate parameter, a length scale, and a parameter that describes the anisotropy of the flow. Values for these parameters can be estimated by using high-frequency sonic anemometer data and can also be approximated from lidar data. However, the Mann (1994) turbulence model is technically only valid in the surface layer under neutral conditions and is not valid in complex terrain.

2.3.2 Six-beam method

To reduce variance contamination caused by the DBS and Velocity-Azimuth Display ~~(VAD) scans~~ (VAD; Browning and Wexler, 1968) techniques, Sathe et al. (2015b) proposed a new six-beam scanning technique for Doppler lidars ~~. While the DBS and VAD techniques involve using the radial velocities to estimate the u, v, and w wind components, then calculating the variance , the six-beam technique uses the variances~~ that utilizes the variance of the radial ~~velocities measured at six different beam positions to estimate the variance and covariance components~~velocity. Newman et al. (2016b) tested the six-beam method with a scanning lidar at the Boulder Atmospheric Observatory ~~(BAO)~~ in Erie, Colorado, and compared six-beam variance estimates to estimates from sonic anemometers on a tower at the site. Newman et al. (2016b) found that while the six-beam method did generally reduce variance contamination in comparison to estimates from a WC lidar~~that used the DBS technique~~, errors in the different radial velocity variance estimates caused large errors and even negative values in the resulting u and v variance estimates. Better estimates of the radial velocity variance are likely needed from lidars to obtain accurate results ~~from~~ for the six-beam technique.

2.3.3 Multiple lidars

While single lidars require measurements around a scanning circle to estimate the three-dimensional velocity field, multiple scanning lidars can be pointed toward a particular volume of air to obtain turbulence estimates with much higher spatial resolution (e.g., Calhoun et al., 2006; Fuertes et al., 2014; Newsom et al., 2015; Newman et al., 2016a). To collect turbulence measurements, multi-lidar systems must be temporally and spatially synchronized with a high degree of accuracy. Synchronization techniques have been developed for a set of user-customized scanning lidars (Vasiljevic et al., 2014), but are currently not easily implemented on most other scanning lidars. In addition, scanning lidars are much more expensive than commercially available vertically profiling lidars, particularly if more than one scanning lidar is required for operation.

2.3.4 Structure functions

Structure functions describe the spatial correlation of a variable at different separation distances (e.g., Stull, 2000). If the turbulence is isotropic and the turbulence length scale is large, the structure function can be approximated by the Kolmogorov (1941) model and used to estimate the velocity variance. Krishnamurthy et al. (2011) used scanning lidar data from a field campaign to calculate structure functions in both the along-beam and azimuthal directions and fit the functions to the Kolmogorov (1941) model to obtain estimates of the velocity variance. The lidar data used by Krishnamurthy et al. (2011) were obtained from a series of plan-position indicator (PPI) scans with high azimuthal resolution, which is typically not available from a scanning strategy used by a ~~commercially available~~ vertically profiling lidar.

2.3.5 Doppler spectrum

As discussed by Mann et al. (2010), the spectral density of a particular radial velocity, v_r, is essentially a weighted count of all the positions within the probe volume where the radial velocity is equal to v_r. The weighting occurs because the intensity

of the lidar beam is highest at the center of the probe volume and drops off for distances in either direction from the probe volume center. The ensemble-averaged spectrum can then be related to the probability density function of the radial velocity at each position within the probe volume. Given this relation, the unfiltered ("true") variance can be obtained from the second central moment of the Doppler spectrum. If the lidar is mounted on a turbine nacelle and pointing upstream, as in Branlard et al. (2013), it can be assumed that the wind field is homogeneous along the lidar beam and that the probability density of v_r is approximately uniform along the probe volume. However, if a ground-based, vertically profiling lidar is used, the mean wind field will not be uniform along the lidar's line-of-sight and the effects of shear must be taken into account when estimating the unfiltered variance from the Doppler spectrum (Mann et al., 2010). Currently, this method is more clearly defined for continuous wave lidars, as the Doppler spectra of pulsed lidars are affected by the finite length of the probe volume in addition to turbulent fluctuations.

2.3.6 Summary

Several methods are currently available for obtaining more accurate turbulence estimates from Doppler lidars. Only a few methods were discussed here; a more extensive discussion of turbulence retrieval techniques can be found in Sathe and Mann (2013) and Sathe et al. (2015a). Most of these methods require the fitting of models and the use of very specific scanning strategies that can currently only be achieved with expensive scanning lidars. The Doppler spectrum method is promising for continuous wave lidars, but requires knowledge of the Doppler spectrum obtained at each lidar beam position, which is usually not available in the output of ~~commercially available~~ vertically profiling systems. Thus, there is clearly a need for a turbulence estimation method that can be implemented on ~~commercially available~~ vertically profiling lidars that use DBS and VAD techniques and that does not require high-resolution data from a sonic anemometer. Details of the new turbulence estimation method proposed in this paper are discussed in the next section.

3 ~~Methodology~~TI error model: L-TERRA

The TI error model described in this work, L-TERRA, was initially developed for the WC pulsed Doppler lidar. Future work will involve expanding L-TERRA to different lidar configurations and scanning strategies, although the basic framework for the model will stay the same. The different modules of L-TERRA in its current form are described in this section. ~~In addition, the power prediction model that was used to relate errors in TI to errors in power prediction is introduced.~~

3.1 ~~TI error model: L-TERRA~~

A flowchart depicting different methods for correcting TI with L-TERRA is shown in Fig. 2. L-TERRA contains several modules that reduce the lidar TI error in different ways. For each of the main modules, outlined in red in Fig. 2, ~~many~~ several different methods are available to reduce the TI error. For example, four different methods were evaluated to reduce noise: a spike filter, and three different methods discussed by Lenschow et al. (2000) (Lenschow 1, Lenschow 2, and Lenschow 3). Some methods can only be applied to the u, v, and w velocity data while others can only be applied to radial velocity data,

v_r; thus, two different model paths can be followed for volume averaging and variance contamination, depending on which wind speed parameters are selected to calculate the variance. In this work, only model combinations that use the u, v, and w velocity components were evaluated, as not all vertically profiling lidars include the line-of-sight wind speed in the output files. All possible combinations of ~~different~~ the different u, v, and w methods were tested on each data set to determine which combination produced the largest reduction in TI ~~and power prediction error~~ mean absolute error (MAE).

3.1 Preprocessing

3.1.1 ~~Preprocessing~~

Several steps are taken before the lidar data enter the TI correction process. First, values of u, v, and w are calculated from the raw ~~WC radial velocity~~ lidar time series. For the WC lidar, the wind speed components can be calculated in two different ways: by estimating new u, v, and w components every time the lidar beam moves to a new position (i.e., ~~every~~ just under 1 s) or by estimating a single value of each of the wind components for every 4-s scan, similar to a VAD technique. In Section ~~??~~5, both the 1-s and 4-s techniques are used to calculate the wind components in the evaluation of L-TERRA.

Next, the data are interpolated to a grid with constant temporal spacing, as statistical measures such as the calculation of variance and spectra require that the frequency resolution of the measurements is constant. The mean horizontal wind speed and shear parameter are calculated before L-TERRA is applied, as these parameters are required for implementation of L-TERRA and are relatively unaffected by the errors that plague turbulence measurements.

The 10min mean horizontal wind speed, $\overline{U}$, is defined by Eq. 2:

$$\overline{U} = \overline{(u^2 + v^2)^{1/2}}, \tag{2}$$

where u and v are the east-west and north-south wind components, respectively, and the overbar denotes temporal averaging. The shear parameter, α, is derived from the standard power law equation (International Electrotechnical Commission, 2005):

$$U(z) = U(z_r)\left(\frac{z}{z_r}\right)^{\alpha}, \tag{3}$$

where z is height above ground and z_r is a reference height. Eq. 3 can be simplified by letting $U(z_r)z_r^{-\alpha}$ equal a constant β, as in Clifton et al. (2013). The power law then becomes the following:

$$U(z) = \beta z^{\alpha} \tag{4}$$

A 10min mean value of α can be found by taking the natural logarithm of Eq. 4 and fitting the resulting equation to a straight line. In this work, values of $\overline{U}$ measured by the WC between 40 and 200 m were used to calculate values of α.

The raw wind speeds are rotated into a new coordinate system by forcing $\overline{v}$ and $\overline{w}$ to 0 and aligning u with the 10min mean wind direction (e.g., Kaimal and Finnigan, 1994). The TI is then defined by Eq. 5:

$$TI = \left(\frac{\sigma_u}{\overline{u}}\right) \times 100\%, \tag{5}$$

where σ_u is the standard deviation of u over a 10min period, defined in the new coordinate system, and $\overline{u}$ is the 10min mean wind speed. Eq. 5 gives the initial lidar-estimated value of the horizontal TI. ~~A similar~~ The same procedure was used to calculate TI from the ~~cup and~~ sonic anemometer data used in this work. Cup anemometer TI was calculated using the mean horizontal wind speed and standard deviation in the cup anemometer output data stream. As the main purpose of this work is to bring lidar TI estimates closer to point measurements from any type of reference device on a met tower, the difference in the way TI is calculated for cup and sonic anemometers is not of paramount importance.

3.1.1 ~~Instrument noise~~

3.2 Instrument noise

After the lidar data are processed, different techniques are used to remove noise and spurious data. A standard way to remove outliers from a time series is to use a spike filter (e.g., Vickers and Mahrt, 1997). A basic spike filter was evaluated for the model in addition to several methods developed by Lenschow et al. (2000). These methods involve the use of the lidar's velocity spectrum or autocovariance function to estimate the amount of noise in the variance measurements from the lidar.

3.2.1 ~~Volume averaging~~

3.3 Volume averaging

Two methods were evaluated to mitigate the effects of volume averaging: structure functions and spectral extrapolation. As discussed in Section 2.3.4, structure functions can be estimated using available lidar data and fit to ~~modeled forms of structure functions~~ models to estimate turbulence parameters (e.g., Krishnamurthy et al., 2011). By fitting the lidar data to a model, the reduction of turbulence due to volume averaging is mitigated. Although the estimation of structure functions with a lidar is optimized with the use of a high-resolution PPI scan, as in Krishnamurthy et al. (2011), structure functions can also be estimated from DBS scans. A longitudinal (i.e., along-beam) structure function can be estimated from a DBS scan by using velocity data from different range gates along the same radial. An azimuthal structure function can be estimated by combining data from different azimuthal directions at the same height, although the fit is likely to be poor for a DBS scan due to the small number of azimuthal angles where data are collected.

Another method of mitigating volume averaging is to model the lidar velocity spectrum and use the model to extrapolate the spectrum to higher frequencies. The high-frequency part of the modeled spectrum can then be integrated to obtain an estimate of the variance that is not measured by the lidar as a result of spatial and temporal resolution (e.g., Hogan et al., 2009). The standard Kaimal spectrum used for wind energy (e.g., Burton et al., 2001) requires three parameters for fitting: the mean wind

speed, the variance, and a length scale. The first two parameters can be calculated from the lidar data directly, while the last parameter must be estimated. In this work, the length scale was estimated in two different ways: by minimizing the difference between the actual velocity spectrum and the Kaimal spectrum (Spectral Fit 1) and by calculating the integral time scale from the raw velocity time series, which can then be related to the integral length scale through Taylor's frozen turbulence hypothesis (Spectral Fit 2).

3.3.1 ~~Variance contamination~~

3.4 Variance contamination

Methods to reduce variance contamination include the six-beam technique developed by Sathe et al. (2015b), discussed in Section 2.3.2, and the use of Taylor's frozen turbulence hypothesis with data from the WC's vertical beam to estimate the change in the vertical velocity across the lidar scanning circle (Newman, 2015). ~~Variance contamination proved to be~~ The six-beam technique can be adapted for the WC DBS scans by estimating the variance from the five different radial beam positions (four off-vertical and one vertical) and solving a system of five equations to determine the variance and co-variance components. Although the covariance of the u and v components cannot be determined with this method, the ~~most difficult effect to capture in L-TERRA. Thus, work with a lidar simulator is currently ongoing to make further refinements to this module of L-TERRA~~three velocity variance components can be estimated (Newman et al., 2016b). Taylor's frozen turbulence hypothesis can be used to correct either the raw wind speed (Taylor 1) or the variance directly (Tayor 2).

3.4.1 ~~Machine learning~~

~~The previous three modules (instrument noise, volume averaging, and variance contamination) constitute the physics-based corrections of L-TERRA. These modules rely only on data from the lidar itself and use theory rather than mathematical models. While the physics-based corrections do reduce lidar TI error, there is still some error in the lidar TI estimates in comparison to estimates from a cup or sonic anemometer. Thus, machine-learning methods were used in the final step of L-TERRA to bring lidar TI estimates even closer to met tower estimates.~~

3.5 Machine learning

~~Three~~ Two machine-learning methods were evaluated as part of L-TERRA: random forests ~~, support vector regression,~~ and multivariate adaptive regression splines (MARS). Random forest models are developed by averaging multiple decision trees that were trained on different subsets of the data. By averaging tens or hundreds of decision trees, the variance of the overall model is reduced significantly (Friedman et al., 2001). Random forests were evaluated because they are relatively easy to understand and have previously been used for wind energy applications (e.g., Clifton et al., 2013; Bulaevskaya et al., 2015). ~~Support vector regression and MARS are both~~ MARS is essentially a stepwise regression model, where different coefficients and basis functions are used to predict the output depending on each region in the dataset (Friedman, 1991). MARS is well-

suited for the prediction of physical processes ~~. More information on random forests, support vector regression, and MARS can be found in , , and , respectively~~due to its ability to model non-linearities and interactions among variables.

Potential predictor variables for the machine-learning models were divided into two broad categories: atmospheric state and lidar operating characteristics. Variables that were evaluated as predictor variables in L-TERRA are given in Table 1. Atmospheric state variables included shear parameter, mean wind speed, Doppler spectral broadening, and u and w velocity variances. Lidar operating characteristics included signal-to-noise ratio (SNR) and internal instrument temperature. ~~Mean wind speed also affects data quality, as lidars cannot measure turbulence at low wind speeds as accurately as a result of relative intensity noise .~~ In all, ~~14~~ 18 predictor variables were considered for the machine learning portion of L-TERRA.

~~Sensitivity of the TI error to the various predictor variables was assessed following the guidelines in Annex L of the new committee draft of IEC 61400-12-1 . Data from the Atmospheric Radiation Measurement (ARM) site, described in Section 4, were used to assess the importance of different predictor variables in predicting lidar TI error. First, predictor variables were binned and bin-means of the TI percent error corresponding to each bin were calculated. A least-squares technique was then used to calculate a regression line between the predictor bin centers and the bin-means of the TI percent error. Sensitivity, defined as the product of the regression line slope and the standard deviation of the predictor variable, was then calculated for each predictor. The sensitivity gives the approximate change in the TI error for a change in the predictor variable that is equivalent to one standard deviation of the variable. All predictor variables had sensitivity values over 0.5, which indicates a significant relationship between all the predictor variables and the TI error, according to Annex L of IEC 61400-12-1 .~~

3.6 Comparison to previous methods

~~As the~~ In contrast to the methods discussed in Sect. 2.2, L-TERRA uses only information that is available from a standard vertically profiling lidar. The physics-based corrections in L-TERRA require only data from the lidar itself, while the machine learning module in L-TERRA can be trained using either cup or sonic anemometer data. The majority of the ~~predictor variables are related to the atmosphere, which is a highly synergistic system, it is likely that one or more of the variables are correlated to one another. Thus, a correlation matrix was calculated for the potential predictor variables. For pairs of variables with a correlation coefficient of over 0.5, the predictor with a lower sensitivity value was removed from the list of potential predictor variables. The final predictor variables were as follows: TI from the physics-based corrections, α, SNR, σ_w^2 (w velocity variance), spectral broadening, instrument internal temperature, and pitch of the lidar.~~

3.7 ~~Power prediction~~

~~Reduction in TI error was related to reduction in turbine power prediction error through the use of a power prediction model. Simulations used in were again used to develop a power prediction model for the 1.5MW WindPACT turbine . First, 3-D flow fields with varying degrees of wind shear and TI were created using TurbSim . These flow fields were then used as input for the turbine simulator FAST to model the response of the WindPACT turbine to flow fields with different degrees of shear and turbulence. The 10min mean hub-height wind speed, hub-height TI, and shear parameter were extracted from the TurbSim output and the 10min mean turbine power was extracted from the FAST output. These parameters, in addition to the turbine~~

operating range, were then used to train a random forest model. The trained model accepts values of mean wind speed, TI, and shear, and predicts the 10min mean power that would be produced by the 1.5MW WindPACT turbine.

Lidars can accurately measure the mean wind speed and the shear parameter, but do not measure the same values of TI as a cup or sonic anemometer (see Section 2.2). Thus, TI is the only variable in the power prediction model that is likely to differ between a lidar and a traditional met tower instrument. In Section **??**, values of $\overline{U}$ corrections in L-TERRA can be implemented with fewer than 20 lines of code, and α from a WC lidar from three different field campaigns are used as inputs to the power prediction model, in addition to values of TI from both the lidar and collocated sonic and cup anemometers. The difference between the lidar- and sonic- or cup-measured TI is then related to errors in the predicted powerthe models employed in L-TERRA are well-documented in the literature and simple to understand. It takes approximately 0.1 seconds to run L-TERRA for a single 10min period, making it easy to implement in real time. As discussed in Sect. 5.1, a stability-dependent version of L-TERRA can be used to adapt to changing conditions and apply corrections appropriate for the current atmospheric stability regime.

4 Data sets

L-TERRA was tested on data from three two different locations: the Southern Great Plains ARM site in Lamont, Oklahoma ; the BAO in Erie, Colorado; (Fig. 3), and an operational wind farm in the Southern Plains region of the United States(Figs. 3–**??**). . The ARM site, a field measurement site operated by the U.S. Department of Energy, contains several remote sensing and in-situ instruments (Mather and Voyles, 2013). From November 2012 to June 2013, a WC lidar owned by Lawrence Livermore National Laboratory was deployed at the ARM site approximately 100 m from a 60m tower. Gill Windmaster Pro 3-D sonic anemometers are mounted on the tower at 25 and 60 m AGL and collect velocity data at a frequency of 10 Hz.

The BAO is a field site located approximately 25 km east of the Rocky Mountain foothills . The WC was deployed at the BAO near a 300m tower from February to April 2014. The 300m tower was instrumented with twelve 3-D sonic anemometers, with two sonics mounted on opposite booms every 50 m from 50 to 300 m AGL. RM Young sonics, owned by the University of Oklahoma, were mounted on the northwest booms of the tower, and Campbell Scientific CSAT3 sonics, owned by the National Center for Atmospheric Research, were mounted on the southeast booms.

The WC was WC was also deployed at an operational wind farm in the Southern Great Plains. (Due to a nondisclosure agreement with the wind farm, we cannot disclose the location or details of the wind farm.) The WC was located on the wind farm from November 2013 to July 2014, with a break from February to April 2014 while the WC was located at the BAOdeployed for a different field experiment. During the wind farm deployments, the WC was sited in the same enclosure as a met tower with standard wind energy instrumentation, including a cup anemometer at the turbine hub height. For the winter deployment, the WC was located near a met tower on the north end of the wind farm, and for the spring/summer deployment, the WC was moved to the a tower enclosure at the south end of the wind farm, in accordance with the dominant wind direction during each season at the wind farm. Data shown here are restricted to wind directions corresponding to turbine inflow.

Although the simulated 1.5MW WindPACT turbine used in this work has a hub height of 84 m, none of the sites were instrumented with cup or sonic anemometers at the 84m measurement height. Thus, the closest height to 80 m that contained both lidar and met tower data at each site was defined as the "hub height" for that site. This resulted in hub heights of 60, 100, and 80 m at Sample scatter plots of met tower versus lidar TI for the ARM site , BAO, and wind farm, respectively.

Histograms of the 10min mean hub-height wind speed, shear parameter, and hub-height SNR from the different sites are shown in Fig. **??**. While the ARM site and Southern Plains wind farm experienced similar atmospheric conditions, conditions at the BAO site were characterized by lower values of the mean wind speed, shear parameter, and SNR. Part of this discrepancy, particularly in SNR, could be due to the different "hub heights" at each site; the SNR values shown for the BAO are from 100 m while the SNR values for the ARM site and wind farm are from 60 4a, and 80 m, respectively. As SNR is strongly tied to aerosol concentrations, which generally decrease with height, SNR is expected to be lower for measurement heights that are further from the ground . However, SNR values from the 50m height at the BAO were also generally lower than SNR values at the ARM site and wind farm. During the BAO campaign, much lower WC data availability was noted in comparison to the ARM site and wind farm campaigns. The lower data availability and SNR values at the BAO can be largely attributed to westerly flow from the direction of the Rocky Mountains, which brings cleaner air with a lower aerosol concentration to areas downwind of the mountains . also noted lower WC lidar data availability in the Colorado foothills region in comparison to a site in Iowa.

TI is shown as a function of mean wind speed and stability class for the different sites in Fig. **??**. corresponding regression statistics for the raw TI are shown in Table 3. Stability classes were stratified according to the value of α measured by the WC between 40 and 200 m, with thresholds given in Table 2. These thresholds are loosely based on the thresholds used in Wharton and Lundquist (2012). The shear parameter α was used as a proxy for stability in this work, as other stability parameters such as the Richardson number and Obukhov length require temperature measurements at different heights, which are not available from a lidar.

The Normal Turbulence Model , indicated by the black lines in Fig. **??**, predicts a sharp decrease in TI as $\overline{U}$ increases and the denominator in Eq. 5 becomes larger. With the exception of some outliers, this trend is largely followed for the TI and $\overline{U}$ values at the different sites. At the ARM siteand wind farm, TI values measured by the lidar are close to those predicted by the NTM under neutral conditions. At these sites, low shear conditions (near-zero or negative shear parameter) tended to be associated with low wind speeds and higher TI values. Low shear conditions often occur when the atmosphere is unstable, resulting in buoyant mixing, a uniform wind speed profile, and higher amounts of turbulence . In contrast, high shear conditions (large positive shear parameter) tended to be associated with higher wind speeds and lower TI values. High shear conditions often occur when the atmosphere is stable; mixing and turbulent motions are inhibited and wind speed tends to decrease with height as frictional effects from the surface become less dominant. At the BAO, low shear conditions were often associated with very low wind speeds and high TI, similar to the other sites. However, the striation of the $\overline{U}$/TI curve by shear parameter is not as prominent at the BAO, and high TI values were often associated with large shear parameters. At this site, the magnitude of TI is likely strongly affected by the low SNR values at the site (Fig. **??**) and complex terrain, in addition to the diurnal heating cycle. Thus, relations between TI, wind speed, and shear parameter are not as clear at the BAO.

Figures ?? and ?? demonstrate the large differences in atmospheric conditions and lidar data quality that can occur in different locations. Thus, the deployment of the same WC lidar at three different sites alongside met towers provides an excellent opportunity to assess the accuracy of lidar-measured TI at different locations. In addition, the large amount of lidar and met tower data collected during the experiments can be used to evaluate the effects of TI error on wind power prediction and to quantify the improvement in power prediction that occurs when lidar TI estimates are improved under different atmospheric conditions.

5 TI error and effects on power prediction

Scatter plots of met tower versus lidar 10min mean wind speed and TI for all three sites are shown in Fig. ??. Mean wind speeds measured by the lidar are extremely close to those measured by the met tower instruments, with regression line slopes near 1 and nearly all coefficient of determination (R^2) values greater than or equal to 0.99. There is slightly more scatter between the lidar and sonic mean wind speeds at the BAO (Fig. ??), likely because SNR values at the BAO were lower and the lidar data quality was not as good in comparison to the other two sites. The excellent comparison of met tower and lidar mean wind speeds indicates that the WC lidar could accurately measure the mean flow at the different sites. However, large discrepancies between the TI measured by the lidar and the met tower instruments were noted at all three sites.

At the ARM site, α was strongly related to the sign of TI errors, with the WC overestimating TI under unstable conditions and underestimating TI under stable conditions (Fig. 4a). The over- and underestimation of TI was likely due to the effects of variance contamination and volume averaging, respectively. Regression line slopes increase with decreasing stability (Table 3), as in Sathe et al. (2011). In this region of the United States, the shear parameter is strongly tied to the atmospheric stability (e.g., Newman and Klein, 2014), likely because the diurnal transition of the atmospheric boundary layer largely controls the wind speed profile in flat terrain (e.g., Arya, 2001). Initial TI error trends from the wind farm data set are remarkably quite similar to those found in the ARM data set (Fig. ??c). This is not surprising, as both data sets were collected in a similar region with similar terrain and diurnal transitions.

Table 4).

In contrast, lidar TI errors at the BAO did not follow a distinct pattern according to the shear parameter (Fig. ??). Flow in this area is affected by complex terrain in addition to diurnal trends, so the shear parameter is likely not an accurate indicator of the atmospheric stability. At the BAO, nearly all the lidar TI measurements were overestimates in comparison to the sonic anemometers. As SNR values at the site were generally much lower in comparison to the ARM site (Fig. ??), more noise was likely present in the lidar data at the BAO, resulting in TI overestimates. The lower SNR at the site also contributed to low lidar data availability and a much smaller number of data points in comparison to the other two sites.

5 L-TERRA results

Next, lidar and met tower TI data were used as inputs for the power prediction model, in addition to

5.1 Application of physics-based corrections

First, data from each site were examined individually to assess the performance of L-TERRA. For both the ARM site and the wind farm, all possible combinations of the physics-based corrections described in Section 3 were evaluated. Initially, the ~~hub-height wind speed, shear parameter, and turbine operating range. As discussed in Section ??, power was predicted for the 1.5MW WindPACT turbine, which has a hub height of 84 m and a rated wind speed of 11.5~~ model combination that produced the lowest overall TI MAE was selected as the optimal model combination for that particular site. Data were filtered to avoid mast shadowing, and 10min periods where the mean wind speed was less than 4 m s^{-1} ~~. Power error is expressed as a function of lidar TI error in Fig. ??, with different colored circles corresponding to different parts of the power curve. At all three sites,~~ were not used to evaluate L-TERRA, as the standards outlined in IEC 61400-12-1 (International Electrotechnical Commission, 2013) restrict remote sensing classification to wind speeds between 4 and 16 m s^{-1}.

The optimal model combination was the same for both sites and is shown in the ~~largest power errors were found in the wind speed region just above and below the rated wind speed, where power sensitivity to turbulence is highest (Fig. ??). In this region, even small TI errors of 1%–2% can result in power errors above 2.5% of the rated power. This trend is evident at all three sites, although there are relatively fewer points in this transition region at the BAO as a result of the lower mean wind speeds experienced at~~ first row of Table 5. Several slightly different model combinations produced similar MAE values at both sites, suggesting that there may actually be multiple optimal combinations of L-TERRA at each site when the MAE is minimized. It may be useful to consider other parameters in determining the optimal model combination, such as regression line statistics or sensitivity of TI error to atmospheric stability. However, minimizing the MAE is a standard approach for determining optimal model combinations and provides a useful baseline combination for evaluating L-TERRA. Application of this initial L-TERRA model combination resulted in a modest reduction of lidar TI MAE from 1.5% to 1.4% at the ARM site and from 1.48% to 1.39% at the wind farm (not shown).

By examining the change in lidar TI after each step in L-TERRA, it was determined that some corrections decreased error under stable conditions while increasing error under unstable conditions, and vice versa. This is not surprising, as the magnitude and sign of TI errors was strongly dependent on atmospheric stability at both sites (Tables 3, 4) as a result of the different factors that affect TI error under different stability conditions. Thus, optimal model combinations were next determined separately for the ~~BAO in comparison to the other two sites (Fig. ??). The largest TI errors occur at lower wind speeds (near 75% of the rated wind speed). However, because power sensitivity to TI is low in this region of the power curve, these large TI errors did not often translate to large errors in predicted power. For wind speeds above the rated wind speed, power error increases steadily with increasing TI error, but most power errors are below 0.5% of the rated power.~~three different bulk stability classes to form a stability-dependent version of L-TERRA (L-TERRA-S). Optimal model combinations were very similar for both sites and are shown in Table 5.

~~Power percent error is expressed as a function of shear parameter and WC TI in Fig. ??. Atmospheric conditions range from low-shear, high TI environments at the upper left corner of the plots (i.e., unstable conditions) to high-shear, low-TI environments at the lower right corner of the plots (i.e., stable~~ For all three stability classes, a spike filter was the optimal noise

removal technique. Only the model chain for stable conditions ~~). At all three sites, power errors are negligible for wind speeds near 75% of the rated wind speed and are generally smaller than 0.5% for wind speeds around 125% of the rated wind speed. The largest errors , as also shown in Fig. ??, occur for wind speeds near rated. (Note the small number of colored boxes in Fig. ?? for wind speeds near and above the rated wind speed. Colored boxes are only shown for bins with three or more data points, and wind speeds at the BAO were generally quite low, as previously discussed.) For the ARM site and the wind farm, the sign of the errors changes when moving from stable to unstable conditions , with power overestimates occurring~~ included a volume averaging correction, likely because volume averaging effects on TI are largest under stable conditions~~and power underestimates occurring under unstable conditions. As the WC typically underestimates TI under stable conditions (Fig. ??), power predictions made with these TI estimates underestimate the effects of TI on power (Fig. 1). Thus, power predictions made under stable conditions with the WCTI values are overestimates of~~ . For unstable conditions, using the velocity time series from the VAD technique produced the largest reduction in MAE. While the raw output time series from the WC is available at 1 Hz (Section 2.1), the ~~true power. In contrast, the WC overestimates TI under unstable conditions (Fig. ??) and thus overestimates the effect of TI on power. In the region near the rated wind speed, overestimating the TI results in predicting a lower amount of power than what is truly produced (Fig. 1). At all three sites, the largest power errors tend to occur under low-shear, high TI conditions, which typically correspond to unstable conditions. This is not surprising, as WC TI estimatesunder unstable conditionshad the largest errors in comparison to met tower measurements, as evidenced by the large slopes~~ VAD technique is typically applied once per full scan to derive the three-dimensional wind vector. For the WC, this results in an output data frequency of 0.25 Hz for the VAD technique. The lower temporal resolution of the VAD technique likely served to artificially reduce some of the effects of variance contamination, as smaller scales of turbulence were not measured.

Scatter plots of ARM site TI data after L-TERRA-S was applied are shown in Fig. 4b and corresponding regression statistics are shown in Table 3. L-TERRA-S served to bring the majority of WC TI estimates closer to the one-to-one line, resulting in regression line slopes of 1.00, 1.01, and ~~lower~~ 1.00 for stable, neutral, and unstable conditions, respectively. In addition, the overall TI MAE decreased from 1.5% to 1.24%. However, R^2 values for neutral and unstable conditions decreased slightly. Thus, although L-TERRA-S improved the accuracy of most lidar TI estimates, it also increased scatter for neutral and unstable conditions.

Results for the wind farm were similar, with overall MAE decreasing from 1.48% to 1.20% and regression line slopes becoming 1.00, 1.04, and 1.05 for stable, neutral, and unstable conditions, respectively (Table 4). R^2 values ~~shown in Fig. ??~~ for all three stability classes decreased slightly.

6 ~~Model results~~

~~In this section, initial results from~~ While the regression line slopes improved significantly with the application of L-TERRA-S, the reduction in MAE was still relatively modest for both sites, and a large amount of scatter still existed. This scatter indicates that the current physics-based corrections in L-TERRA ~~are shown for data from the ARM site, the BAO, and the Southern Plains wind farm. For this initial investigation of~~ do not fully capture all the factors that affect lidar TI error. In the next

section, machine-learning techniques are evaluated as a potential method to model the ~~performance of L-TERRA, the structure function and six-beam techniques were not evaluated, as these techniques require further refinement for inclusion in the model.~~ remaining physics that impact lidar TI error.

5.1 ~~Single data sets~~

5.1 Application of machine-learning techniques

~~First, data from each site were examined individually to assess model performance. For each site, all possible combinations of the error~~

The physics-based corrections described in ~~Section ?? were evaluated. For the~~ the previous section require only data from the lidar itself and do not require any training data. In contrast, machine-learning ~~module, 75% of the data were used for training with the remaining 25% used for testing model performance. The corrected lidar TI was then used as an input parameter for the power prediction model described in Section ??. Ten-minute mean power predicted from the lidar inputs was compared to power predicted from the met towerinputs, and the model combination that produced the lowest power mean absolute error (MAE) was selected as the optimal model combination for that particular site. At all three sites , data were filtered to avoid mast shadowing. In addition, 10min periods where the mean wind speed was less than 3 m s^{-1} were not used to evaluate~~ models must be trained on a dataset before being applied to new data. Typically, a single dataset is split into training and testing datasets in a method known as cross-validation (e.g., Efron and Gong, 1983) to test the accuracy of the model on data that was not included in the training process. As the end goal of L-TERRA ~~, as the WindPACT turbine used in this work has a cut-in wind speed of 3 m s^{-1}.~~

~~The optimal model combinations for all three sites are shown in Table 5. (Note that only one variance contamination option was evaluated for this work, the correction based on Taylor's frozen turbulence hypothesis.) At all three sites, calculating the u~~is to provide accurate lidar TI values at a site that doesn't have a met tower, ~~v, and w wind speed every 1 second using the DBS technique produced better results than calculating new values of the wind speed components every 4 s using data from~~ the ~~entire scan. This is not surprising, as TI calculated from data with higher temporal resolution is more likely to match TI calculated from a cup or sonic anemometer. The noise removal, volume averaging, and~~ machine-learning model in L-TERRA must be trained on one or more sites with a met tower before being applied to a lidar at a new site. Thus, the machine-learning ~~options were slightly different for the different sites. However, the use of corrected TI data in the power prediction model at all three sitessignificantly decreased power MAE values (Table 5).~~ models discussed in this section were trained on the wind farm data and then applied to data from the ARM site for validation.

~~Although L-TERRA improved lidar TI estimates and power predictions, the model does not perform uniformly under all atmospheric conditions. The change in power percent error resulting from application of L-TERRA is shown as a function of shear exponent and~~ To determine appropriate predictor variables for the machine-learning module, a sensitivity analysis was conducted for the WC TI ~~in Fig. ?? for the ARM site and Southern Plains wind farm (results from the BAO are not shown in this figure due to the smaller number of data points at the BAO). L-TERRA generally improves power estimates slightly for~~

wind speeds above and below the rated wind speed, although the error increased for some periods with a small shear exponent (< 0.1). The largest decreases in power percent error occur for low TI values and higher shear exponent values at the ARM site (Fig. **??**), which typically correspond to stable conditions. The power percent error tended to increase for low shear (e.g., unstable conditions) at both sites, but particularly at the wind farmerror at both sites.

Scatter plots of WC versus met tower TI both before and after L-TERRA has been applied to Sensitivity of the lidar TI error to the various predictor variables in Table 1 was quantified following the guidelines in Annex L of the new committee draft of IEC 61400-12-1 (International Electrotechnical Commission, 2013). First, predictor variables were binned and bin-means of the TI percent error corresponding to each bin were calculated. A least-squares technique was then used to calculate a regression line between the predictor bin centers and the test set bin-means of the TI percent error. Sensitivity, defined as the product of the regression line slope and the standard deviation of the predictor variable, was then calculated for each predictor. The sensitivity gives the approximate change in the TI error for a change in the predictor variable that is equivalent to one standard deviation of the variable.

Sample plots showing the response of TI percent error to different variables at the ARM site are shown in Fig. **??**. Data points corresponding to an absolute power percent error greater than 1% are highlighted. Several periods with power percent error greater than 1% are associated with high-shear5. Raw WC TI error was extremely sensitive to the four variables depicted in Fig. 5, with larger TI percent errors for lower wind speeds and SNR values, low-TI conditions (stable). Although L-TERRA improves these TI estimates, bringing them very close to the one-to-one line (Figs. **??**, **??**), these periods are still associated with large power percent errors . This likely occurs because at both of these sites, low-TI, high-shear conditions often correspond to higher wind speeds near the rated wind speed of the 1.5MW WindPACT turbine (Figs. **??**, **??**), where turbulence sensitivity is highest (Fig. **??**). Other large power errors are associated with unstable and neutral conditions with higher TI values. Although the largest TI errors occur under low-shear, high-TI conditions (Figs. 4a, **??**c, **??**, **??**), these large TI errors do not often result in large power errors, as they are usually associated with low wind speeds at both the ARM site and the wind farm (Figs. **??**, **??**), where turbulence sensitivity is low (Fig. **??**). Several of the neutral and unstable TI estimates that are associated with large power errors are initially located above the one-to-one line in Figs. **??** and **??**; these overestimates are likely a result of variance contamination, which is most prominent under unstable conditions . The variance contamination module in L-TERRA reduces these TI overestimates somewhat, but several of the TI estimates associated with large power errors still lie above the one-to-one line and TI errors changing sign from negative to positive for decreasing shear parameter values and increasing raw TI values. After L-TERRA-S was applied, TI error sensitivity to shear parameter and raw lidar TI decreased significantly. This decrease in sensitivity demonstrates a major advantage of L-TERRA-S, as it implies that lidar TI error is no longer strongly dependent on atmospheric stability. However, while L-TERRA-S decreased sensitivity slightly for mean wind speed and SNR, a high dependence of TI error on these two variables is still apparent in Figs. **??** and **??**. Some of the TI values measured by the WC under unstable conditions are initially quite accurate (e.g., the TI estimate lying on the one-to-one line near TI = 20% in Fig. **??**), but decrease as a result of the variance contamination module and become less accurate (Fig. **??**)5c and 5d.

Overall, the six variables for the ARM site with the highest sensitivity values after application of L-TERRA-S were as follows: integral time scale (vertical), SNR, integral time scale (horizontal), mean wind speed, corrected TI, and shear

parameter. (For highly correlated variables, the variable with a higher sensitivity was retained in the list.) These six variables were then used to train a random forest with the wind farm data.

5.2 ~~Combined data sets~~

~~In the previous section, the machine-learning model in L-TERRA was trained and tested at the same site and significantly reduced TI~~Results from application of the trained random forest on the ARM site L-TERRA-S TI values are shown in Fig. 6a. Application of the random forest increased MAE values from 0.77% to 0.91% for stable conditions and from 1.48% to 1.56% for neutral conditions, and ~~power errors at all three sites. However, the goal of L-TERRA is to provide more accurate TI estimates from a stand-alone lidar at a site that does not have a met tower. Thus, the performance of L-TERRA at a site where the machine-learning model was not trained is of paramount importance.~~

~~In this section, the optimal physical corrections (Table 5) were applied to the data from each site. Data from each site were then divided into training and testing data sets, with 75% of the data used for training and 25% of the data used for testing, similar to the procedure used in Section **??**. Different combinations of the training data sets were then used to train the MARS machine-learning model, and the trained model was tested on the test data sets from all three sites. The power MAE for each combination of training and testing sets is shown in Table **??**, with bold values corresponding to model performance when the MARS model is both trained and tested at the same site.~~

~~For the ARM site, training the MARS model with the BAO and/or wind farm data always resulted in higher power MAE values~~ decreased the MAE from 1.58% to 1.55% for unstable conditions in comparison to ~~only training the model with data from the ARM site, although these MAE values were lower than the original MAE before L-TERRA was applied (Table 5). However, training the model with multiple data sets improved the performance of L-TERRA at the ARM site. When the BAO or wind farm training data sets were used individually, the power MAE for the ARM site was 2.13 kW. Training the model with data from both the BAO and~~ the ~~wind farm reduced the power MAE to 2.05 kW. By including data from different sites, the model can be trained on a larger variety of conditions and is thus more likely to perform better at a site where it has not been trained. Results for the BAO are similar , with lower power MAE values produced when both the ARM site and wind farm data are used to train the MARS model.~~ results from L-TERRA-S. For all three stability classifications, R^2 values dropped significantly and a positive bias was induced for low TI values.

~~When the MARS model was trained using data from~~ To determine the cause of this positive bias, the sensitivity values from both sites were compared for the six input variables. While the regression line slope and sensitivity values for the vertical integral time scale, SNR, mean wind speed, and shear were very similar at both sites, sensitivity values for the horizontal integral time scale and corrected TI differed substantially. In particular, the sensitivity of TI error to the corrected TI was 5.51% at the ARM site and ~~tested on~~ four times larger at 22.3% at the wind farm~~site, a large MAE value of 5.14 kW was produced (Table **??**). This large increase in MAE is surprising, as the atmospheric conditions at the ARM site and the wind farm are quite similar (Figs. **??**, **??**). However, when the MARS model parameters were tuned to the ARM site data~~. After removal of the horizontal integral time scale and the corrected TI from the input parameter list, the bias at low TI values largely disappeared (Fig. 6b), suggesting that the positive bias was related to the different sensitivities associated with two of the input

parameters. However, even with these two parameters removed from the input parameter list, the MAE ~~value for the wind farm was reduced from 5.14 kW to 1.51 kW. Optimal tuning parameters for the different machine-learning models will be further investigated in the future when more data sets have been collected~~values still increased in comparison to L-TERRA-S while R^2 values decreased. Results from the MARS model were similar to those from the random forest models.

~~It should also be noted that while the MARS~~ These results highlight two major limitations of using machine-learning ~~model performed best for the ARM and BAO sites, a random forest model produced the lowest power MAE value for the wind farm (Table 5) . Thus, different combinations of~~ techniques to improve lidar TI accuracy in L-TERRA: 1) The most significant input parameters can change from one site to another and will not be known a priori for a new site and 2) Sensitivity of TI error to different input variables depends on the training site and the particular lidar and reference measurements used. To investigate the effect of the training ~~and test sets were also evaluated with the random forest model. The use of a random forest instead of the MARS model reduced the wind farm MAE to 1.59 kW when the ARM training data set was used. However, using~~ dataset used, a random forest ~~also increased the wind farm MAE to 11.95 kW when only the BAO training data set was used to train the model. Using~~ was also trained on 75% of the ARM site data and then applied to the remaining 25%. Training and testing a random forest ~~increased the ARM site MAE to 5.84 kW when the BAO training data set was used. Thus, although the random forest model seems to generally perform better at the wind farm, large errors occur when only the BAO data set is used to train the random forest. It is possible that a random forest does not perform as well as the MARS model when the training data set is associated with different conditions than the testing data set~~ on data from the same site did decrease MAE values in comparison to results from L-TERRA-S, but R^2 values still decreased slightly for neutral and unstable conditions. More importantly, using this technique would preclude L-TERRA from being applied at a new site that doesn't have a met tower.

Although machine learning can be a useful tool for turbine power prediction (e.g., Clifton et al., 2013), it does not appear to be an ideal technique for correcting lidar TI error. Thus, next steps in the development of L-TERRA will involve further refining the physics-based corrections in L-TERRA-S to improve TI estimates in a more robust manner. Rather than relying on modeled patterns, physics-based corrections directly relate lidar measurements to TI errors and substantially improved the accuracy of lidar TI estimates at both sites evaluated in this work (Fig. ~~??~~)~~or when the training data set is small. Histograms of training and testing input parameters indicated that several input parameters from the ARM and wind farm sites were outside the range of parameters used to train the model with the BAO data (not shown). This indicates the importance of using a training data set that encompasses a large range of atmospheric conditions.~~ 4, Tables 3 and 4). However, the current physics-based corrections do not completely eliminate TI error, indicating that the physics that cause TI error are not being entirely captured in L-TERRA-S. Future work will involve the development of a lidar uncertainty framework that outlines all possible causes of lidar error. Different parts of the framework could then be quantified through the use of a simulated flow field and virtual lidar, as in Lundquist et al. (2015).

~~In general, the machine-learning module in L-TERRA can be trained and tested at different sites without a significant increase in power MAE. This is an important finding, as it implies that L-TERRA can be trained at one or more sites and then applied to lidar data at a new site to improve TI estimates. While the MARS model performed well for the ARM and BAO data sets regardless of which training data sets were used,~~ Results from the sensitivity analysis conducted in this section will greatly

assist in determining areas of focus for the lidar uncertainty framework. For example, TI error at both sites was extremely sensitive to the ~~random forest method was generally more well suited to the wind farm data set. Future research will focus on the optimal machine-learning method to use for a particular combination of training and testing data sets~~integral time scale of the w wind component, which is a proxy for the dominant temporal scale of turbulent eddies in the vertical direction. Thus, lidar TI error appears to be strongly affected by the scales of vertical motion present in the area enclosed by the lidar scanning circle, which will contribute to the degree of variance contamination. Currently, no physical models exist that could account for these effects, and so we suggest that this could be a fruitful research direction. In future work, the virtual lidar tool will be used to examine how changes in the vertical flow field across the lidar scanning circle impact TI estimates and how information from the lidar can be used to approximate and remove these effects.

6 Conclusions and future work

Lidars are currently being considered as a viable replacement to meteorological towers in the wind energy industry. Unlike met towers, lidars can be easily deployed at different locations and are capable of collecting wind speed measurements at heights spanning the entire turbine rotor disk. However, lidars measure different values of TI than a cup or sonic anemometer, and this uncertainty in lidar TI estimates is a major barrier to the adoption of lidars for wind resource assessment and power performance testing. In this work, a lidar turbulence error reduction model, L-TERRA, was developed and tested on WC lidar data from ~~three~~ two different sites. The model incorporates both physics-based corrections and machine-learning techniques to improve lidar TI estimates.

~~Main findings from the~~ The main findings from this work can be summarized as follows:

- ~~For mean wind speeds near the rated wind speed of a turbine, small errors of 1%–2% in TI can result in large errors in power prediction~~The difference between TI measured by a cup or sonic anemometer and that measured by a vertically-profiling lidar can be reduced using appropriate physical models of the lidar measurements.
- Performance of L-TERRA improves ~~TI estimates and reduces power MAE at all three sites, although the optimal model configuration depends on the site.~~ substantially when different model configurations are used for different stability conditions (i.e., in L-TERRA-S).
- ~~L-TERRA performs most poorly under high-shear conditions , when mean wind speeds tend to be near the rated wind speed, and under low-shear conditions, when variance contamination can significantly increase lidar TI estimates~~In addition to reducing MAE, L-TERRA-S also reduces sensitivity of lidar TI error to atmospheric stability.
- The accuracy of a machine-learning ~~module in L-TERRA generally reduces power MAE even when the machine-learning model is trained at one site and tested at another site. A larger reduction in MAE occurs when the machine-learning model is trained on more than one data set.~~

  ~~Although the combination of physics-based corrections and machine learning, as implemented in L-TERRA, is a promising method for reducing~~ method in L-TERRA-S is highly dependent on the sensitivity of the lidar TI error ~~, further refinements~~

must be made to different modules in the model to improve performance under a variety of conditions. Future work will include the use to the input parameters at both the training and testing sites.

Further improvements to L-TERRA-S can be made through a better understanding of how atmospheric conditions and lidar operating characteristics impact TI error. This understanding can be achieved through development of a lidar simulator to improve the volume averaging and variance contamination corrections in L-TERRA. Additional data sets will be collected for training and uncertainty framework and testing of the modelframework with modeled atmospheric data. Future work on L-TERRA-S will also include testing with additional datasets, including data sets from complex terrain and different areas of the world. The current version Practical applications of L-TERRA was developed specifically for the vertically profiling WC lidar, but plans are underway to expand L-TERRA to different lidar models and scanning configurationsfor site assessment and power prediction will also be explored in future work.

The development of L-TERRA and other TI correction techniques has significant implications for the wind energy industry, which has traditionally relied on data from fixed met towers. L-TERRA can be applied to commercially available vertically profiling lidars that are commonly used in the wind industry, thus expanding the use of lidars for wind energy applications. Lidars with improved TI estimates can be used for wake studies, site classification, power curve testing, site monitoring, and resource assessment. Improved lidar TI estimates could also help wind energy developers make more informed decisions about turbine selection and wind farm layout. The use of lidars in place of met towers for wind energy applications should allow for more rapid development of wind in regions where it is difficult or costly to install met towers, and the improvement of lidar turbulence estimates will greatly assist in the adoption of lidars in the wind industry.

Acknowledgements. The authors would like to thank the staff of the Southern Great Plains ARM site , the BAO, and the Southern Plains wind farm , Tim Bonin, Bruce Bartram, and Daniel Wolfe of NOAA/ESRL, Lucas Root and Andreas Muschinski from NorthWest Research Associates, and Shiril Tichkule from the University of Colorado at Boulder for assisting with the lidar field deployments. Sebastien Biraud and Marc Fischer of Lawrence Berkeley National Laboratory supplied sonic anemometer data for the ARM siteand Andreas Muschinski of NorthWest Research Associates provided sonic anemometer data for the BAO. Sonia Wharton of Lawrence Livermore National Laboratory provided the WINDCUBE lidar used in this work and assisted with field deployments. Leosphere and NRG Systems provided technical support for the WINDCUBE lidar during the experiments. Conversations with Caleb Phillips at NREL greatly enhanced our understanding of different machine learning models. The ARM Climate Research Facility is a U.S. Department of Energy Office of Science user facility sponsored by the Office of Biological and Environmental Research. This work was supported by the U.S. Department of Energy under Contract No. DE-AC36-08GO28308 with the National Renewable Energy Laboratory. Funding for the work was provided by the DOE Office of Energy Efficiency and Renewable Energy, Wind and Water Power Technologies Office.

References

Arya, S. P.: Introduction to Micrometeorology, vol. 79 of *International Geophysics Series*, Academic Press, Cornwall, UK, 2nd edn., 2001.

Barthelmie, R. J., Crippa, P., Wang, H., Smith, C. M., Krishnamurthy, R., Choukulkar, A., Calhoun, R., Valyou, D., Marzocca, P., Matthiesen, D., Brown, G., and Pryor, S. C.: 3D wind and turbulence characteristics of the atmospheric boundary layer, Bull. Amer. Meteor. Soc., 95, 743–756, doi:10.1175/BAMS-D-12-00111.1, 2013.

Boquet, M., Callard, P., Deve, N., and Osler, E.: Return on investment of a lidar remote sensing device, DEWI Magazine, 37, 2010.

Branlard, E., Pedersen, A. T., Mann, J., Angelou, N., Fischer, A., Mikkelsen, T., Harris, M., Slinger, C., and Montes, B. F.: Retrieving wind statistics from average spectrum of continuous-wave lidar, Atmos. Meas. Tech., 6, 1673–1683, doi:10.5194/amt-6-1673-2013, 2013.

Browning, K. A. and Wexler, R.: The determination of kinematic properties of a wind field using Doppler radar, J. Appl. Meteor., 7, 105–113, doi:10.1175/1520-0450(1968)007<0105:TDOKPO>2.0.CO;2, 1968.

Bulaevskaya, V., Wharton, S., Clifton, A., Qualley, G., and Miller, W. O.: Wind power curve modeling in complex terrain using statistical models, Journal of Renewable and Sustainable Energy, 7, 013 103, doi:10.1063/1.4904430, 2015.

Burton, T., Sharpe, D., Jenkins, N., and Bossanyi, E.: Wind Energy Handbook, John Wiley & Sons, Ltd., 2001.

Calhoun, R., Heap, R., Princevac, M., Newsom, R., Fernando, H., and Ligon, D.: Virtual towers using coherent Doppler lidar during the Joint Urban 2003 Dispersion Experiment, J. Appl. Meteor., 45, 1116–1126, doi:10.1175/JAM2391.1, 2006.

Chang, W. S.: Principles of Lasers and Optics, Cambridge University Press, 2005.

Clifton, A., Kilcher, L., Lundquist, J. K., and Fleming, P.: Using machine learning to predict wind turbine power output, Environmental Research Letters, 8, 024 009, doi:10.1088/1748-9326/8/2/024009, 2013.

Clifton, A., Boquet, M., Roziers, E. B. D., Westerhellweg, A., Hofsäß, M., Klaas, T., Vogstad, K., Clive, P., Harris, M., Wylie, S., Osler, E., Banta, B., Choukulkar, A., Lundquist, J., and Aitken, M.: Remote sensing of complex flows by Doppler wind lidar: Issues and preliminary recommendations, Tech. Rep. NREL/TP-5000-64634, NREL, http://www.nrel.gov/docs/fy16osti/64634.pdf, 2015.

Efron, B. and Gong, G.: A leisurely look at the bootstrap, the jackknife, and cross-validation, The American Statistician, 37, 36–48

Emeis, S.: Measurement Methods in Atmospheric Sciences: In-Situ and Remote, Borntraeger Science Publishers, 257 pp., 2010.

Friedman, J., Hastie, T., and Tibshirani, R.: The Elements of Statistical Learning, vol. 1, Springer Series in Statistics Springer, Berlin, 2001.

Friedman, J. H.: Multivariate adaptive regression splines, The Annals of Statistics, pp. 1–67, 1991.

Fuertes, F. C., Iungo, G. V., and Porté-Agel, F.: 3D turbulence measurements using three synchronous wind lidars: Validation against sonic anemometry, J. Atmos. Oceanic Technol., 31, 1549–1556, doi:10.1175/JTECH-D-13-00206.1, 2014.

Hogan, R. J., Grant, A. L., Illingworth, A. J., Pearson, G. N., and O'Connor, E. J.: Vertical velocity variance and skewness in clear and cloud-topped boundary layers as revealed by Doppler lidar, Quart. J. Roy. Meteor. Soc., 135, 635–643, doi:10.1002/qj.413, 2009.

Huffaker, R. M. and Hardesty, R. M.: Remote sensing of atmospheric wind velocities using solid-state and CO_2 coherent laser systems, Proceedings of the IEEE, 84, 181–204, 1996.

International Electrotechnical Commission: Wind turbines - Part 1: Design requirements, Tech. rep., IEC 61400-1, Geneva, Switzerland, 2005.

International Electrotechnical Commission: Wind turbines - Part 12-1: Power performance measurements of electricity producing wind turbines, Tech. Rep. Committee draft ed., IEC 61400-12-1, Geneva, Switzerland, 2013.

Jonkman, J. M. and Buhl Jr., M. L.: FAST user's guide, Tech. Rep. No. NREL/EL-500-38230, National Renewable Energy Laboratory, Golden, CO, 2005.

Kaimal, J. and Finnigan, J.: Atmospheric Boundary Layer Flows: Their Structure and Measurement, Oxford University Press, 1994.

Kolmogorov, A. N.: The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers, Doklady AN SSSR, 30, 301–304, 1941.

Krishnamurthy, R., Calhoun, R., Billings, B., and Doyle, J.: Wind turbulence estimates in a valley by coherent Doppler lidar, Meteorological Applications, 18, 361–371, doi:10.1002/met.263, 2011.

Krishnamurthy, R., Choukulkar, A., Calhoun, R., Fine, J., Oliver, A., and Barr, K.: Coherent Doppler lidar for wind farm characterization, Wind Energy, 16, 189–206, doi:10.1002/we.539, 2013.

Lenschow, D. H., Wulfmeyer, V., and Senff, C.: Measuring second-through fourth-order moments in noisy data, J. Atmos. Oceanic Technol., 17, 1330–1347, doi:10.1175/1520-0426(2000)017<1330:MSTFOM>2.0.CO;2, 2000.

Lundquist, J. K., Churchfield, M. J., Lee, S., and Clifton, A.: Quantifying error of lidar and sodar Doppler beam swinging measurements of wind turbine wakes using computational fluid dynamics, Atmos. Meas. Tech., 8, 907–920, doi:10.5194/amt-8-907-2015, 2015.

Malcolm, D. and Hansen, A.: WindPACT turbine rotor design study, Subcontract rep. No. SR-500-32495, National Renewable Energy Laboratory, Golden, CO, 2006.

Mann, J.: The spatial structure of neutral atmospheric surface-layer turbulence, J. Fluid Mech., 273, 141–168, doi:10.1017/S0022112094001886, 1994.

Mann, J., Peña, A., Bingöl, F., Wagner, R., and Courtney, M. S.: Lidar scanning of momentum flux in and above the atmospheric surface layer, J. Atmos. Oceanic Technol., 27, 959–976, doi:10.1175/2010JTECHA1389.1, 2010.

Mather, J. H. and Voyles, J. W.: The ARM Climate Research Facility: A review of structure and capabilities, Bull. Amer. Meteor. Soc., 94, 377–392, doi:10.1175/BAMS-D-11-00218.1, 2013.

Newman, J. F.: Optimizing lidar scanning strategies for wind energy turbulence measurements, Ph.D. thesis, University of Oklahoma, Norman, Oklahoma, USA, 2015.

Newman, J. F. and Klein, P. M.: The impacts of atmospheric stability on the accuracy of wind speed extrapolation methods, Resources, 3, 81–105, doi:10.3390/resources3010081, 2014.

Newman, J. F., Bonin, T. A., Klein, P. M., Wharton, S., and Newsom, R. K.: Testing and validation of multi-lidar scanning strategies for wind energy applications, Wind Energy, doi:10.1002/we.1978, 2016a.

Newman, J. F., Klein, P. M., Wharton, S., Sathe, A., Bonin, T. A., Chilson, P. B., and Muschinski, A.: Evaluation of three lidar scanning strategies for turbulence measurements, Atmos. Meas. Tech., 9, 1993–2013, doi:10.5194/amt-9-1993-2016, 2016b.

Newsom, R. K., Berg, L. K., Shaw, W. J., and Fischer, M. L.: Turbine-scale wind field measurements using dual-Doppler lidar, Wind Energy, 18, 219–235, doi:10.1002/we.1691, 2015.

Peña, A., Hasager, C. B., Gryning, S.-E., Courtney, M., Antoniou, I., and Mikkelsen, T.: Offshore wind profiling using light detection and ranging measurements, Wind Energy, 12, 105–124, doi:10.1002/we.283, 2009.

Sathe, A. and Mann, J.: A review of turbulence measurements using ground-based wind lidars, Atmos. Meas. Tech., 6, 3147–3167, doi:10.5194/amt-6-3147-2013, 2013.

Sathe, A., Mann, J., Gottschall, J., and Courtney, M. S.: Can wind lidars measure turbulence?, J. Atmos. Oceanic Technol., 28, 853–868, doi:10.1175/JTECH-D-10-05004.1, 2011.

Sathe, A., Mann, J., Barlas, T., Bierbooms, W. A. A. M., and van Bussel, G. J. W.: Influence of atmospheric stability on wind turbine loads, Wind Energy, 16, 1013–1032, doi:10.1002/we.1528, 2013.

Sathe, A., Banta, R., Pauscher, L., Vogstad, K., Schlipf, D., and Wylie, S.: Estimating turbulence statistics and parameters from ground- and nacelle-based lidar measurements: IEA Wind expert report, DTU Wind Energy, Denmark, 2015a.

Sathe, A., Mann, J., Vasiljevic, N., and Lea, G.: A six-beam method to measure turbulence statistics using ground-based wind lidars, Atmos. Meas. Tech., 8, 729–740, doi:10.5194/amt-8-729-2015, 2015b.

Sjöholm, M., Mikkelsen, T., Mann, J., Enevoldsen, K., and Courtney, M.: Time series analysis of continuous-wave coherent Doppler lidar wind measurements, IOP Conference Series: Earth and Environmental Science, 1, 012 051, doi:10.1088/1755-1307/1/1/012051, 2008.

Sjöholm, M., Mikkelsen, T., Mann, J., Enevoldsen, K., and Courtney, M.: Spatial averaging-effects on turbulence measured by a continuous-wave coherent lidar, Meteor. Z., 18, 281–287, doi:10.1127/0941-2948/2009/0379, 2009.

Slinger, C. and Harris, M.: Introduction to continuous-wave Doppler lidar, in: Summer School in Remote Sensing for Wind Energy, Boulder, CO, 2012.

Sonnenschein, C. M. and Horrigan, F. A.: Signal-to-noise relationships for coaxial systems that heterodyne backscatter from the atmosphere, Appl. Opt., 10, 1600–1604, doi:10.1364/AO.10.001600, 1971.

Stawiarski, C., Träumner, K., Knigge, C., and Calhoun, R.: Scopes and challenges of dual-Doppler lidar wind measurements — An error analysis, J. Atmos. Oceanic Technol., 30, 2044–2062, doi:10.1175/JTECH-D-12-00244.1, 2013.

Strauch, R. G., Merritt, D. A., Moran, K. P., Earnshaw, K. B., and De Kamp, D. V.: The Colorado wind-profiling network, J. Atmos. Oceanic Technol., 1, 37–49, doi:10.1175/1520-0426(1984)001<0037:TCWPN>2.0.CO;2, 1984.

Stull, R. B.: Meteorology for Scientists and Engineers, Brooks/Cole, 2nd edn., 2000.

Vasiljevic, N., Courtney, M., and Mann, J.: A time-space synchronization of coherent Doppler scanning lidars for 3D measurements of wind fields, Ph.D. thesis, Denmark Technical University, Denmark, 2014.

Vickers, D. and Mahrt, L.: Quality control and flux sampling problems for tower and aircraft data, J. Atmos. Oceanic Technol., 14, 512–526, doi:10.1175/1520-0426(1997)014<0512:QCAFSP>2.0.CO;2, 1997.

Wagner, R., Antoniou, I., Pedersen, S. M., Courtney, M. S., and Jørgensen, H. E.: The influence of the wind speed profile on wind turbine performance measurements, Wind Energy, 12, 348–362, doi:10.1002/we.297, 2009.

Wainwright, C. E., Stepanian, P. M., Chilson, P. B., Palmer, R. D., Fedorovich, E., and Gibbs, J. A.: A time series sodar simulator based on large-eddy simulation, J. Atmos. Oceanic Technol., 31, 876–889, doi:10.1175/JTECH-D-13-00161.1, 2014.

Weitkamp, C.: Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere, vol. 102 of *Springer Series in Optical Sciences*, Springer Science & Business Media, 2005.

Wharton, S. and Lundquist, J. K.: Assessing atmospheric stability and its impacts on rotor-disk wind characteristics at an onshore wind farm, Wind Energy, 15, 525–546, doi:10.1002/we.483, 2012.

Wharton, S., Newman, J. F., Qualley, G., and Miller, W. O.: Measuring turbine inflow with vertically-profiling lidar in complex terrain, J. Wind Eng. Ind. Aerodyn., 142, 217–231, doi:10.1016/j.jweia.2015.03.023, 2015.

Wyngaard, J. C.: The effects of probe-induced flow distortion on atmospheric turbulence measurements, J. Appl. Meteor., 20, 784–794, 1981.

Potential Predictor Variables	
Atmospheric State	**Lidar Operating Characteristics**
~~Corrected~~ Original TI	SNR
~~σ_u^2~~ Corrected TI	Instrument pitch
~~σ_w^2~~ σ_u^2	Instrument roll
~~$\overline{U}$~~ σ_w^2	Instrument internal temperature
$\overline{U}$	
α	
Horizontal wind speed dispersion	
Vertical wind speed dispersion	
~~Difference between lidar TI and IEC TI model~~ Spectral broadening	
~~Wind direction~~ Integral time scale (horizontal)	
Integral time scale (vertical)	
Stationarity (e.g., Vickers and Mahrt, 1997)	
Maximum instantaneous value of w	
Precipitation	

Table 1. Potential predictor variables evaluated in the machine-learning module of L-TERRA.

Stability Classification	**Shear Parameter Range**
~~Strongly stable $\alpha \geq 0.3$~~ Stable	~~$0.2 \leq \alpha < 0.3$~~ $\alpha \geq 0.2$
Neutral	$0.1 \leq \alpha < 0.2$
Unstable	~~$0 \leq \alpha < 0.1$~~ $\alpha < 0.1$
~~Strongly unstable $\alpha < 0$~~	

Table 2. Stability classifications used in this work.

	MAE		***Slope***		***R^2***	
	Raw	**L-TERRA-S**	**Raw**	**L-TERRA-S**	**Raw**	**L-TERRA-S**
Stable	1.02	0.77	0.90	1.00	0.88	0.89
Neutral	1.48	1.48	1.04	1.01	0.89	0.87
Unstable	1.95	1.58	1.12	1.00	0.79	0.77
All	1.50	1.24	1.05	1.00	0.86	0.86

Table 3. Mean absolute error (MAE) and slope and R^2 values of regression lines for WC TI compared to met tower TI before and after L-TERRA-S has been applied.

	MAE		*Slope*		R^2	
	Raw	**L-TERRA-S**	**Raw**	**L-TERRA-S**	**Raw**	**L-TERRA-S**
Stable	1.13	0.89	0.91	1.00	0.92	0.91
Neutral	1.26	1.17	1.07	1.04	0.86	0.85
Unstable	1.96	1.53	1.14	1.05	0.85	0.83
All	1.48	1.20	1.07	1.04	0.88	0.89

Table 4. As in Table 3, but for Southern Plains wind farm data.

Stability Classification	Wind Speed	Noise	Volume Averaging	Variance Contamination
All	Raw	Lenschow 1	None	Taylor 1
Stable	Raw	Spike filter	Spectral Fit 2	Taylor 1
Neutral	Raw	Spike filter	None	Taylor 2
Unstable	VAD	Spike filter	None	Taylor 1

Table 5. L-TERRA model combinations that minimized TI MAE for different stability conditions at the ARM site and Southern Plains wind farm.

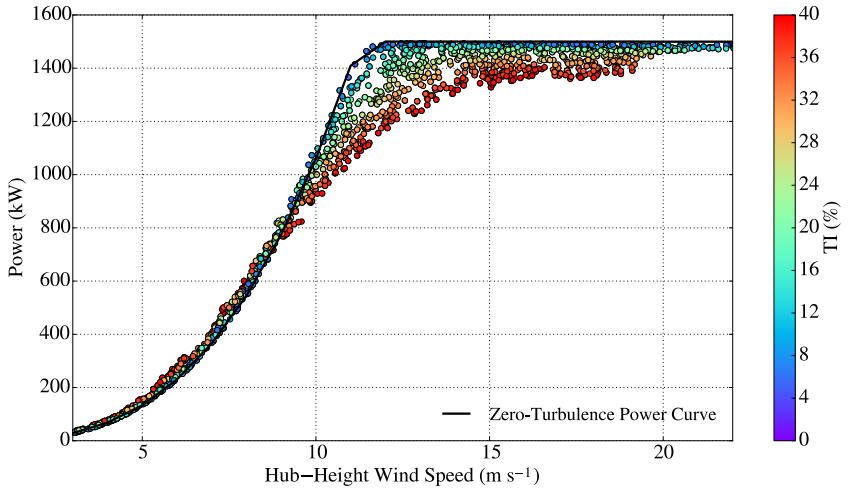


Figure 1. ~~a)~~ Ten-minute mean power for the 1.5MW WindPACT turbine (Malcolm and Hansen, 2006) as a function of mean hub-height wind speed and TI. Mean power is derived from FAST (Fatigue, Aerodynamics, Structures, and Turbulence) simulations (Jonkman and Buhl Jr., 2005). Power curve where there is no turbulence in the flow ("zero-turbulence power curve") is shown for reference. Rated wind speed is 11.5 m s^{-1}.~~b) Sensitivity of power to TI as a function of mean hub-height wind speed for FAST simulations. Only data where the shear exponent is approximately equal to 0.2 are shown. Sensitivity was approximated as the regression line slope for power versus TI in different wind speed bins.~~

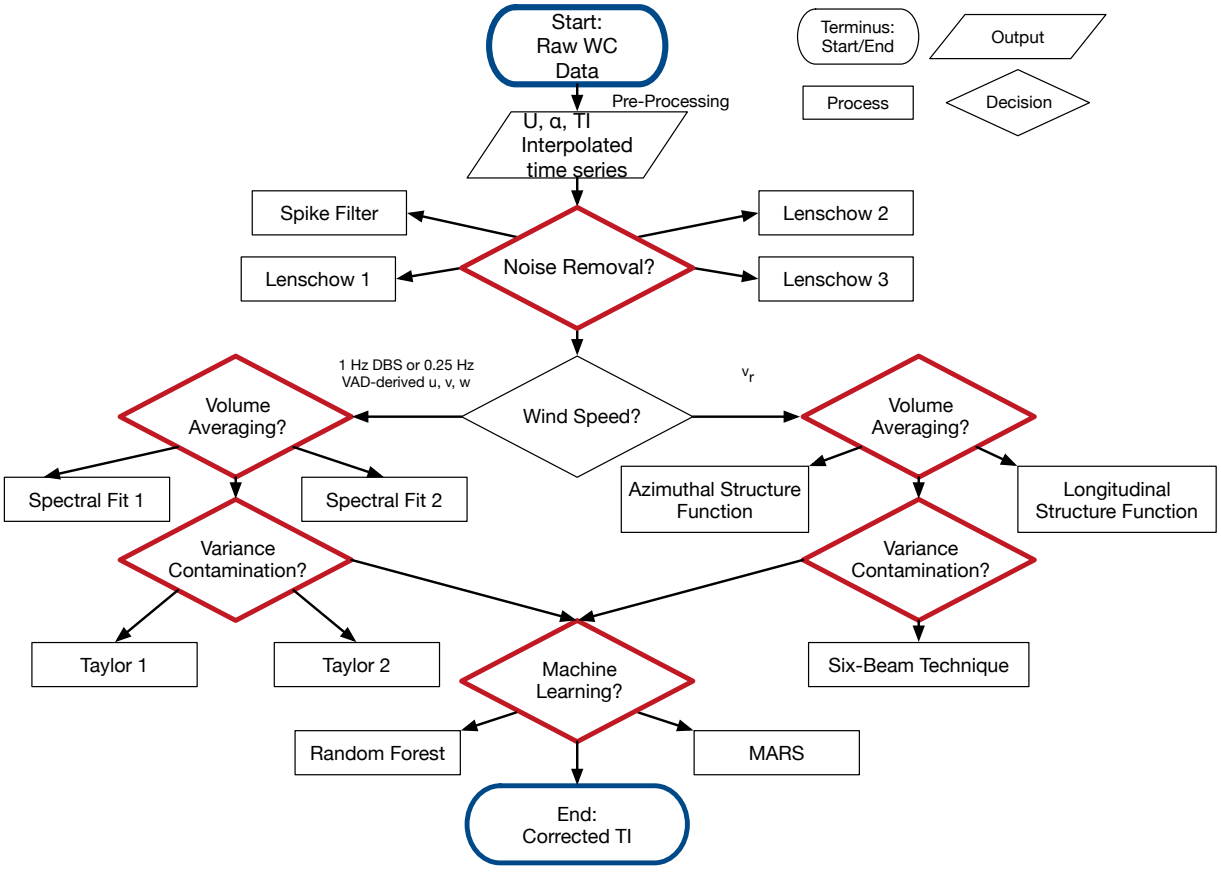


Figure 2. Flowchart depicting different methods for correcting TI with L-TERRA. Starting and ending points are indicated by blue-outlined ovals and modules are indicated by red-outlined diamonds.

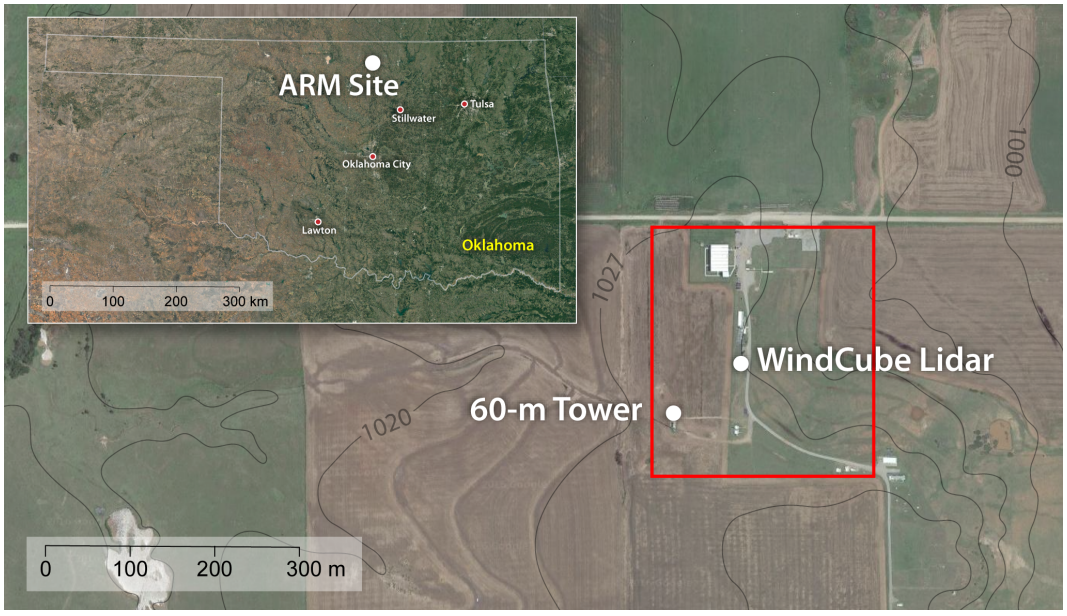
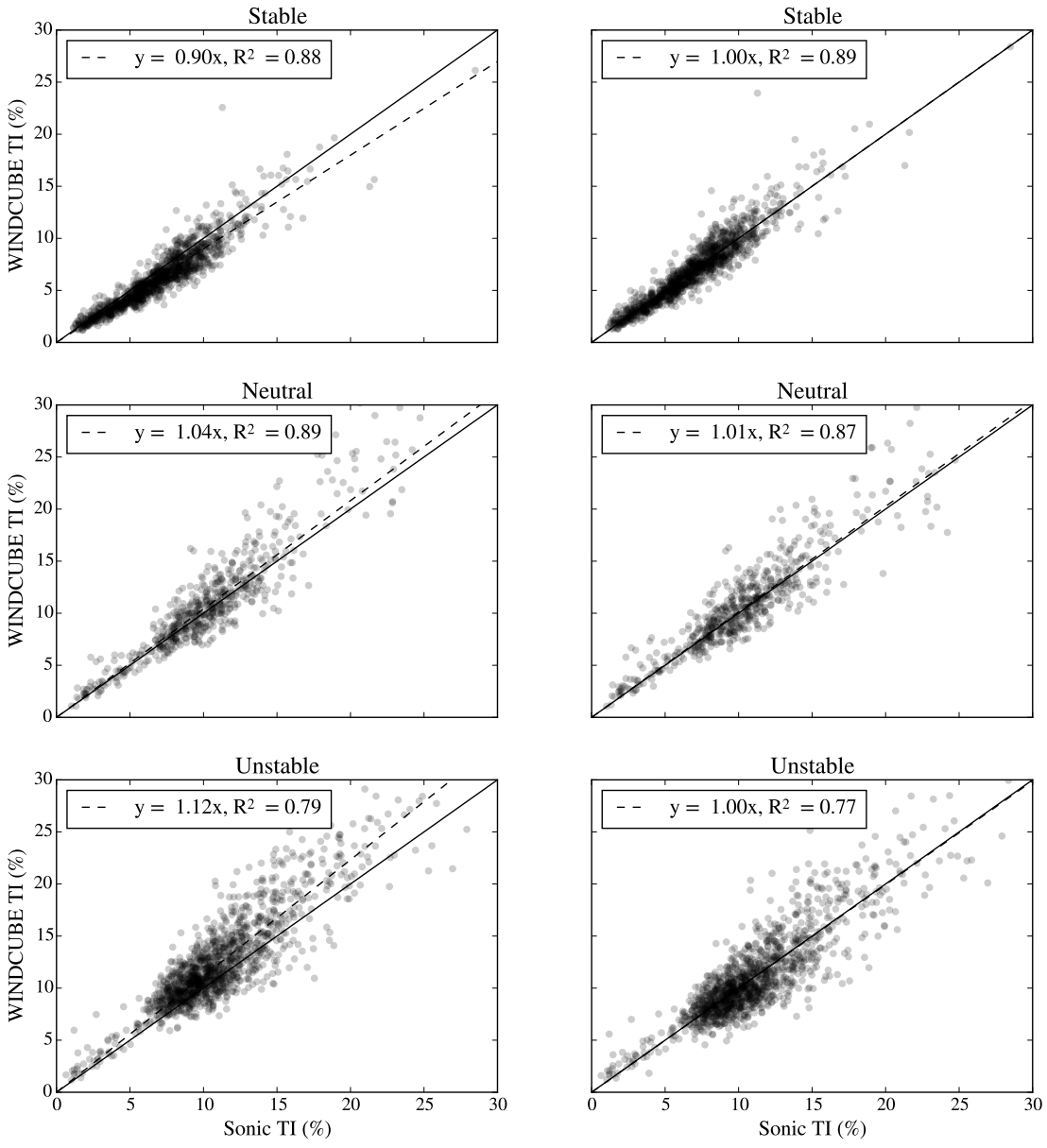


Figure 3. Inset: Google Earth image of the state of Oklahoma. Location of Southern Great Plains ARM site is denoted by white marker. Larger figure: Google Earth image of the central facility of the Southern Great Plains ARM site (outlined in red box) with overlaid elevation contours in feet. Elevation map is from the United States Geological Survey and uses contour intervals of approximately 10 feet (3.05 m). Locations of WC lidar and 60m tower are indicated by white markers.



(a) Raw

(b) With L-TERRA-S

Figure 4. Scatter plots of met tower vs. WC TI for data from 60m measurement height at the ARM site a) before and b) after L-TERRA-S has been applied. One-to-one line and regression lines are shown for reference and regression line statistics are shown in figure legends.

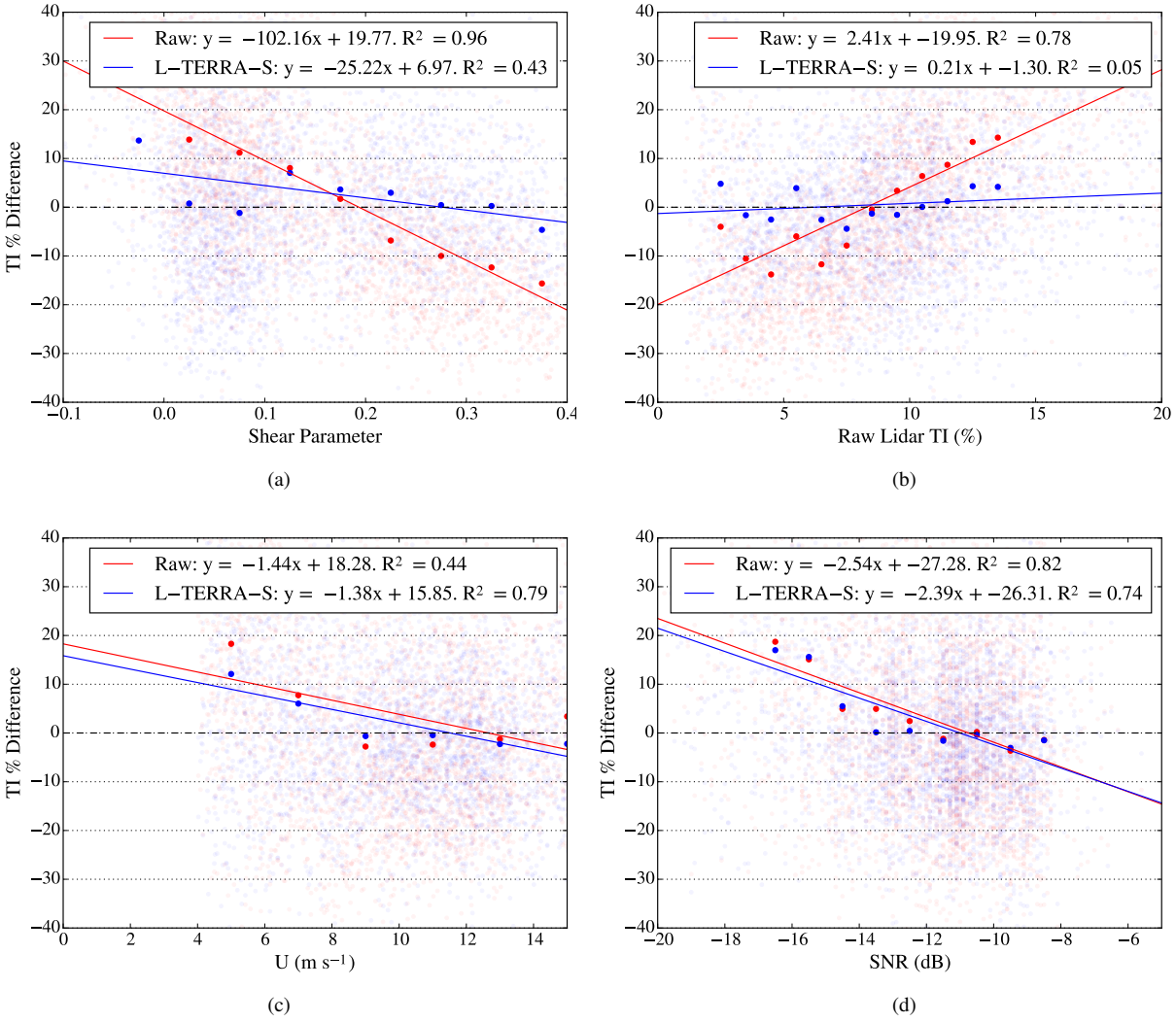
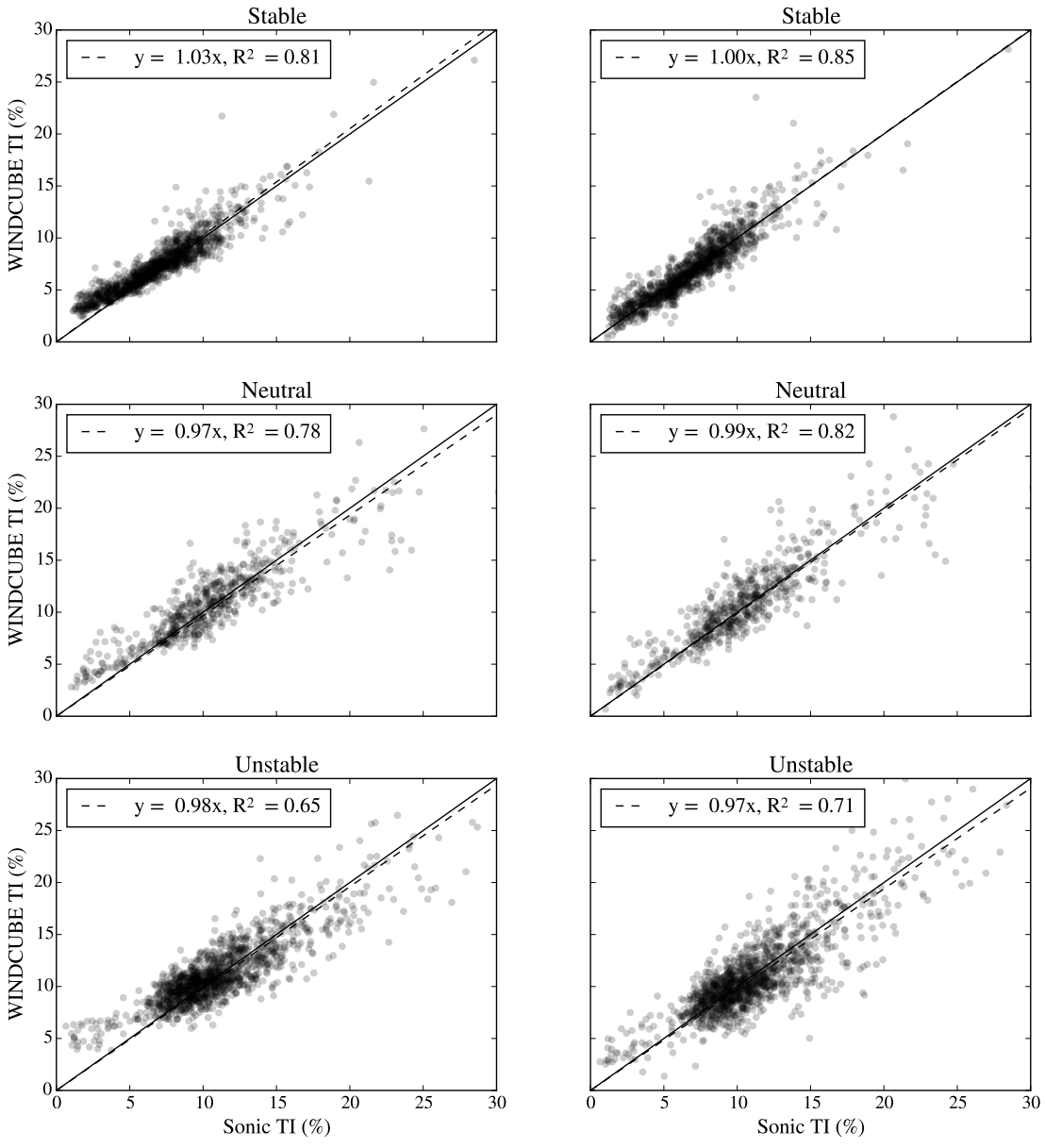


Figure 5. Percent difference between WC and sonic TI for the ARM site as a function of a) shear parameter b) raw WC TI c) mean wind speed and d) SNR. Differences are shown both before (red circles) and after (blue circles) L-TERRA-S has been applied. Solid circles correspond to averages of binned data and solid lines correspond to regression line fits to bin-means, following the procedure in Annex L of IEC 61400-12-1, Draft Edition (International Electrotechnical Commission, 2013).



(a) Random Forest 1 (six input variables)

(b) Random Forest 2 (four input variables)

Figure 6. Scatter plots of met tower vs. WC TI for data from 60m measurement height at the ARM site a) after application of the first random forest described in the text and b) after application of the second random forest. One-to-one line and regression lines are shown for reference.