Long-term uncertainty quantification in WRF-modeled offshore wind resource off the US Atlantic coast

by

N. Bodini and S. Castagneri

The paper nicely brings together methods and simulations from previously peer-reviewed papers by the first author to answer the question of uncertainty quantification connected to offshore long-term modelled hub-height wind speed in the case of limited (only near surface) observational data. The method considers modelling uncertainty as well as observationally based uncertainties connected to near surface observations uncertainties and height extrapolation uncertainties using a peer-reviewed machine learning approach.

General comment

I find the paper interesting and providing new and original material. It is clearly within the scope of WES and introduces a novel approach to uncertainty quantification. By backtracking references to previous papers, the methods and simulations are in principle repeatable. However, some more information on the machine learning algorithm such as choice of hyperparameters etc. would be beneficial for better reproducibility.

The paper is written in a clear and well-structured way. It would, however, benefit from a clearer statement of objectives in the final part of the introduction.
As errors are depending on averages in time and space it should also be more clearly emphasized in both the abstract, introduction and conclusions that results are for hourly average windspeeds.

A final general remark is that most results are given in terms of rmse (m/s) and for readers unfamiliar with the climatology of the region in is difficult to assess if uncertainties in the order of 2-3 m/s are large or small. Some indication of relative uncertainties or typical hour-to-hour wind speed variability would help alleviate this.

Specific comments

- Abstract: In the abstract there is only one sentence on the results. A slightly more elaborate description of the results (such as how errors are connected to stability etc.) would be beneficial.

- Introduction: The introduction provides a thorough review of uncertainties in NWP modelling, but relatively little is given on the observational uncertainties ranging from instrument uncertainties to representation uncertainty. A possible starting point could be the below review of forecasting errors.
  https://www.sciencedirect.com/science/article/pii/S1364032122004221

- Introduction: The final part of the introduction would benefit from a clearer statement of objectives and sub-objectives to make it clear for the reader what the objectives are.

- Section 2.1: The paper references to a previous paper for details about the numerical data. This is fine, but I would have preferred that information such as horizontal resolution, number of levels and output frequency is given in the text and not only stated in the table.

- Section 2.2.1: The authors list the different lidar data used and state that "proprietary quality checks" are done. A few sentences or a reference to what the proprietary quality check contains would be useful.

- Section 2.2.1: I found it unclear what was done in terms of averaging for the lidar data. The authors state in the section 2.2.1 that "*We kept only hourly time stamps where 140 m wind speed, near-surface wind speed, near-surface wind direction, air temperature, and sea surface temperature were all available*." Does this mean that when relations between near surface and 140m winds are estimated based on 5 or 10 min. values, or are there some kind of hourly averaging done before the relations are estimated? If the later is the case please not how missing data is treated in the averaging.

- Section 2.2.2: How is missing data treated in the averaging the NDBC buoy observations?

- Section 3.1: This describes the Machine learning algorithm for wind speed vertical extrapolation. The input are hourly averaged NDBC buoy observations, but it is unclear if the lidar data used for learning is 10 min or hourly averaged lidar data?

- Section 3.1: The authors perform a 5-fold cross validation using a consecutive 20% of the observations. The test data should be unconnected to the training data for a realistic estimation of a possible training to testing degradation of the quality. As windspeeds may easily have lag correlations, the authors should justify that the test data indeed is unconnected to the training data. Any remaining relations between the training and test data could seriously hamper the error statistics. A possible test would be to see if splitting the data chronologically (i.e. 20% of the years used for testing) gives a similar result as the selected validation method.
- The authors should also state both the training and testing validation in order to discuss possible overfitting issues.

- Section 3.1: The authors test different hyperparameters, but the final choices of parameters are not stated.

- Section 3.2: In the total uncertainty estimates the uncertainty components are added in quadrature (equation 1). This way of estimating the total errors is only valid if the errors are not correlated. Please add an analysis showing that the error components are uncorrelated to justify equation 1.

- Section 4: The authors state that "*This comparison clearly confirms how an NWP model's boundary condition and parametric uncertainty, which can be quantified from the variability across a numerical ensemble, can only quantify a limited component of the full model uncertainty*", but how is the ensemble uncertainty quantified? The 16 member ensemble will be highly correlated and this have to be reflected in the ensemble uncertainty estimation through the error covariance.

- Section 5: This consist of the Machine learning wind speed vertical extrapolation validation. A key question that pops up in the round-robin validation is how correlated the sites are. The results It is not easy to understand if the slight increase in rmse of around 12.5% for the round-robin validation compared to the same-site validation is high or low, as long as there is no analysis of among-site correlation.

- Section 5: Figure 6 gives the quantification of the relative importance of the various input variables. It is not stated, but I guess this is the random forest out-of-bag estimate? If so, this analysis requires that the features are uncorrelated, so this has to be established before any clear conclusions can be drawn from the analysis. If they are correlated, this should be stated with the possible influences this may have on the results.

- Section 5: The relative importance analysis (fig 6) indicates that only two features account for most of the predictability. If this is the case, one could argue that the machine learning model, could be heavily simplified. This could be investigated, by looking at the quality degradation between a model using all features and one using only the two main features. Using for example AIC (Aike Information Criterion) or other measures that penalizes unnecessary complicated models, one could get an idea if having all features is worth it compared to a simpler model.

- Section 5: It is unclear what the benefit of the random forest model is compared to a simple more transparent multilinear regression model? Is the added complexity worth it? It would be very interesting if the authors could provide an estimate of the added value of the non-linear random forest model compared to a simple multilinear regression model with only two features.

- Section 6: The author states that uncertainty increases as the distance from the lidars, used to train the machine learning model, increases. But this is not incorporated in the extrapolation uncertainty. Instead it may turn up as part of the WRF uncertainty as the model is compared to data which have gradually lower quality as the validation is moving away from the lidar used in extrapolation training. This questions the attribution of uncertainty to the different terms (wrf vs total). A discussion on this would help the readers to reflect on how the different uncertainty terms should be interpreted and the fact that they might not be independent from each other.

- The conclusion would benefit from some critical discussion on the choice of methods, possible limitations, and outstanding research questions.

Figures

- Figure 1: State the difference between diamonds and dots in the caption.

- Figure 3: Consider skipping this.

- Figure 5 and 7: The scatter plot caption needs a sentence about the color coding indicating density of the data. State that the data shown is hourly averaged wind speeds.

- Figure 5 and 7: Plotting some quantile-quantile values as dots on top of the current plot would make it easier to see any systematic differences in the quantiles.