Manuscript WES-2021-35

Response to Anonymous Referee 1

Dear Editor, dear Anonymous Referee,

First of all, we would like to thank the Anonymous Referee for the objective high-quality review of the manuscript. We think it will greatly improve the quality of the paper and we are grateful for the time spent on it.

We will provide below some answer to the Anonymous Referee remarks, before producing a revised version of the paper.

MAJOR COMMENTS:

1. In general, the main limitation of the current draft is the lack of scientific rigor in the presentation of the approaches used in the analysis. Please remember that you want to make your work replicable after one has read your paper. Below I have added several specific comments to provide examples of this issue. Please, re-consider your technical and statistical explanations and add details where needed. Also, the use of the word "error" and some other qualitative terms should be carefully revised, too.

Indeed, the work should be replicable, therefore, we will give more details on the machine learning in the revised version, including the full parametrization. This problem is not really a lack a scientific rigor, but rather than we did not give enough details. Moreover, we will define clearly in the beginning the statistical terms that we use, like 'errors' (error time series) and 'biais' (systematic error, mean error, MBE). We will also change several labels of the y-axis of some key figure that were confusing: we will use 'wind speed bias in %' instead of 'mean wind speed percentage of error'.

Due to ATMOSKY corporate constraints, we cannot provide the full parametrization of the WRF model. We think this is not a limitation since the numerical modelling is not the main purpose of this paper. It is just used as source of meteorological parameters related to the atmospheric stability. We will only provide the references that were used to choose the PBL scheme.

2. The error quantification in Section 4.2 comes after the ML-derived wind speed distributions have been corrected based on hub-height observations in Section 4.1. If I got this right, this step does not make much sense from a scientific point of view. You are presenting error metrics for the ML-extrapolated winds after using hub-height observations to correct them: first, why should one even consider applying a ML-based approach to derive hub-height winds if observations are available? And then, the ML error quantification should NOT involve any correction using hub-height observations, as this would of course "artificially" improve the ML performance.

It is of course not scientifically correct to use the Lidar data to do an 'a posteriori' correction, as the Anonymous Referee explains. However, it does not affect the validation of the machine learning algorithm itself. Indeed, the machine learning was designed to provide instantaneous wind speeds,

(not wind statistics) and was validated with independent data in Section 3, before the 'a posteriori' correction presented in section 4.2.

We did this 'a posteriori' correction because, in the framework of wind energy, we are interested in the wind distribution, and we noticed that the tails of the distribution were not reproduced correctly after the machine learning. Since this problem has intrinsic causes related to the sensor and the algorithm (i.e., the satellite does not detect low wind speeds and the random forest algorithm cannot extrapolate), we considered it was acceptable to do this 'a posteriori' correction since we expected it to be stable and the same everywhere.

However, as the Anonymous Referee mentions, this is can be seen as a weak point of the study. Therefore, we did a further analysis and found that this effect could be compensated by applying a specific linear model on the values in the distribution tails, instead of the random forest algorithm. In other words, we now combine two models: the random forest for typical values and a linear model for extreme values. This improvement will be presented in the revised paper and removes the need for an 'a posteriori' correction.

SPECIFIC COMMENTS:

1. Line 27: "floating lidars" are not a "method" to retrieve hub-height wind speed. Please rephrase.

ok

2. L. 33: what type of "heterogeneities" are you referring to in the offshore wind context?

The term "heterogeneities" is indeed not appropriate. We are referring to the limited capability to resolve small scale phenomena that the numerical model cannot 'see' due to the scale truncation, and therefore flattens. We propose to rephrase it as follows:

"Conversely, numerical models provide outputs over the entire area of interest, but they tend to flatten extremes and are not capable of resolving small scale phenomena due to their physics and resolution. "

3. Please add references to the first and second paragraphs of the introduction (currently, the first paragraph has zero).

The beginning of the first paragraph corresponds to general statements regarding common practices in the wind energy industry that we observe. We will add this reference for example: "Best Practices for the Validation of U.S. Offshore Wind Resource Models Mike Optis, Nicola Bodini, Mithu Debnath, and Paula Doubrawa, National Renewable Energy Laboratory."

Concerning the end of the first paragraph, fact that the uncertainties related to numerical models are unknown is also a very general statement. Even when a validation of numerical models with in-situ instruments is available, these results cannot theoretically be applied to other locations or periods, which leads the investors to look for real data. We do not have a specific reference for this limitation of numerical models, but we think it is well-known truth.

4. L. 38: a lidar IS a remote sensing instrument, too. Please rephrase to better convey your comparison between lidars and satellite-based retrievals.

We will rephrase this sentence more clearly to explain that lidars are remote sensing instruments.

However, from the point of view of this study, we prefer to call lidars simply 'in-situ' instruments in the rest of the paper, rather than 'remote sensing' instruments. This is due to the scales. For example, in the case of rain, a disdrometer is a remote sensing instrument using a beam, but we would say it is an in-situ when compared with a meteorological radar, because of the important difference of scales involved. Here, it is the same between the satellite and the Lidars.

5. Lines 65-70: please specify in which section you deal with each of the tasks you mention in the paragraphs.

ok

6. Line 80: wasn't WRF developed at NCAR? Add a reference.

ok

7. Please add a reference for GFS, too.

ok

8. Did you run WRF in chunks (e.g., separate runs every month) or as a single run?

The WRF was run in chunks of 24h.

9. Figure 1: please make the fonts bigger (in the axes and the colorbar). Also, specify in the caption what the considered time period is. Also, please specify if latitude is N or S, and if longitude is E or W.

ok

10.Section 2.1: which planetary boundary layer scheme did you use in WRF? Please clarify.

We used a specific planetary boundary layer scheme adapted to coastal environments. Although, as explained previously, we cannot provide a full list of the parameters, we will provide references that were used to do it, like:

M. Diallo, Y. Roustanand, E. Dupont. Extrapolating wind field at hub height from synthetic aperture radar (SAR) and ensemble of weather and research forecast (WRF) wind estimates. WindEurope Offshore 2019, 26-28 November 2019, Copenhague, Danemark

11.Throughout the paper, please consider using a different color scheme. The rainbow color has quite a lot of issues, see for example https://www.climate-labbook.ac.uk/2014/end-of-the-rainbow/

Ok, we will use more appropriate colobar, like nipy_spectral.

12.Line 93: again, lidars are not in-situ instruments.

See answer to point 4.

13.L. 95: please provide more details on the QC – is there a reference?

There is no reference. We will provide more details. For example, the minimum number of packets and availability percentage.

14.L. 100: if two lidars were not used in the analysis, I guess there is no need to include them in the map and in the text.

ok

15.Table 1: "lowest" and "highest" instead of "first" and "last" when talking about altitudes.

ok

16.Figure 2: typo on the y-axis label. Also, what do the shading of the dots represent?

The shading of the dote represent the density of dots that is sometimes difficult to see on scatter plots

17.Line 115: what do you mean by "correct on average"? Please provide quantitative assessments.

Ok. 'Correct on average' means that is the exponent of the power law is unbiased when considering all data (average error is 0). However, we decided to refine it depending on the atmospheric stability because it can fluctuate importantly between stable and unstable conditions. We could rephrase by saying: "correct on average but does not distinguishes between the different stability regimes: neutral, stable and instable".

18.L. 147: what do you mean by "around 5 AM or 5 PM"? Please be specific and provide a plot/histogram/table if necessary.

We will give details: "The satellites pass at 5 AM or 5 PM depending on the orbit orientation, descending or ascending, respectively. The exact acquisition time can also vary by plus or minus 30 mn depending on the incidence angle under which the region of interest is observed."

We do not think a histogram would be necessary here.

19.I don't think Figure 3 is needed. In the text, you can simply state that you use data from 2017 to 2019 because the satellite constellation was not fully operational before that. No need to insert an histogram here (if you really want, please move it to the supplementary information).

Ok we will suppress figure 3.

20.What is the difference between the map in Figure 1 and what is shown in Figures 4 and 5? If it is just a matter of the years being plotted, I would suggest plotting the 2017-19 data in Figure 1 directly, so that Figures 4 and 5 can then be removed from the paper as they do not add much information.

Ok. Yes, it would make the paper lighter.

21.Figure 6: "number" instead of "nb". Also in the caption, please specify which methods you are using to provide more context.

ok

22.L. 199: how do you quantify the "error"? What does "around" mean? Please be specific: you want to make your work replicable!

The ‘error’ L.199 relates to the ‘mean wind speed’ calculated in the previous sentence. We will just use ‘bias’ instead in the revised version and not ‘mean wind speed error’.

23.Figure 7: please specify the height at which wind speed is considered (y-axis label). Also, is time on the x-axis UTC or rather local time? Please specify.

Ok. The time is UTC (which often equate to local time here because we are close to Greenwich).

24.L. 214: please provide context as well as references for the sentence “In addition, numerical model outputs are not as reliable as in-situ data, especially in coastal areas.”. Also, what do you mean here with in-situ data? Once again, lidars are not.

Again, see the answer to comment 4 that explains why in this context we used the term ‘in-situ’ instrument for lidars. This makes it easier for the reader to clearly identify the satellite when we use term ‘remote sensing’ and avoid unnecessary confusion. As stated before, we will mention that Lidar are remote sensing instrument in the introduction and explain our choice.

Regarding numerical models, there is no debate that in-situ real data are more accurate. We do not think a reference is needed, because it is obvious that coastal processes involve smaller scales than the ones in open seas, thus increasing potential errors of the model due to the scale truncation. We will explain it in the revised version.

25.L. 222: from the text, it is not immediately clear the purpose of these two ML models. This becomes clear later in the section, but please state that here too.

ok

26.L.224: you need to define what “error” means for you with a precise statistical metric. Bias? RMSE? R2? Or…?

Error is a term that has different meaning in different sentences. It has a meaning only relatively to a specific value, therefore we cannot define it once for all. Here in particular, there might be a confusion because an ‘s’ is missing. The correct sentence is : “the errors between the SAR and Lidar measurements”, which seems clear to us.

27.L. 225: what do you mean by “such a correlation”? Please provide a threshold or a quantitative measure of what you did.

We did not use a criterion, but a simple visual observation of the scatter plots. We will say it was done with ‘visual inspection’. We evaluate the parameter importance, so we check the accuracy of our choices afterwards and a visual inspection is therefore enough here.

28.L. 226: please clarify what the azimuth angle, the incidence angle, the elevation angle, the backscatter are for a reader not familiar with satellite data retrievals.

Ok.

29.Figures 8, 11, 13: please label the x-axis with actual names (rotated to get enough space), not numbers. 30.L. 239: how was the test set built? Was a random half of the data, or…? At all sites, or…?

Ok. The feature importance is obtained by the attribute `feature_importances_` of the random forest regressor using scikit-learn, after the training phase with 50% of the data (all lidars). We will give these details.

31.L. 240: please provide additional details on the machine learning models and their training. Did you use cross-validation? What hyperparameters did you set? What is the structure of the neural network chosen? Did you train the model at all sites together, or at one site at a time?

The models were trained at all sites together because the Lidars are not very far and located in areas with similar wind patterns. Moreover, since the lidar measurements were obtained during different campaigns with different instruments, we avoid obtaining a result related to instrument differences by doing so.

The hyperparameters were set using cross-validation, we will provide them in the revised version, as well as the neural networks MLP architecture. The hyperparameters found are not very different from the scikit-learn default ones for the random forest regressor.

32.L. 240: are you talking about mean bias? Please clarify.

We use the term 'bias' as it is commonly used, meaning a systematic error (the mean of the errors time series, or mean error). In order to avoid confusion, we will add that it is also sometimes called Mean Bias Error (MBE) by other authors and define it clearly in the beginning.

3**3.Figure 9: again, specify what the different colors for the dots means. Also, larger fonts please. Also, is this for the test set only? At all sites? Please clarify.**

Ok, yes this is the test set. For the colors, again, this is just the density.

34.Figure 10: "error" is too vague. Do you mean bias? Also, I would suggest mobbing this figure to the supplement, as all the information in it is already included in Figure 9, and it is not discussed in detail in the main text.

The term 'errors' used here does not seem vague to us since we explain it is the errors between the SAR and the Lidars. However, we will add that the red dots represent the 'bias' if it is not obvious (the mean of these errors).

We could move this figure to the supplements.

35.Section 3.2: the first paragraph could be moved to the introduction.

ok

36.Section 3.2: once again, more details are needed to fully understand how the machine learning approach was applied. See my other specific comment above for specific questions that need to be addressed.

Here too, we will provide the hyperparameters.

37.Figure 12: the y-axis label is not clear to me.

The label is indeed confusing. The y-axis is the percentage of bias compared to the mean wind speed. We will clarify and use 'wind speed bias (%)'.

38.Do you have any explanations on why for some lidars in Fig. 12 the performance decreases with height, while for some others it actually increases?

Actually, the performance does not increase or decrease with altitude depending on the lidar. As explained just above, the y axis is the bias in %, which can be positive or negative. The dispersion that can be observed when the altitude is increasing is just the error bar of our method that slightly

increases with altitude, as it could be expected. When the altitude increases, the bias of our method for a given lidar can be found in a larger interval around 0.

39.Figure 15: please correct the y-axis label.

Ok (we assume you are talking about Figure 14)

40.Section 4.1: it is not clear to me how the correction is performed. Please provide additional details to make your work replicable.

See the answer to major comment 2. We will not do this correction in the revised paper, but improve the method by merging the outputs of two algorithms depending on the wind speed range.

41.In Section 4.2, you state that "Due to the short distances between the Lidars used in this study, such a validation could not be realized here.". To me, lidars that are about 100 km from each other would still allow for a validation using one for training and another for testing

The maximum distance between the lidars is about 80 km. This would be enough on land, but on the seas, we have reservations because of the homogeneity of the wea surface and of the wind field. Moreover, there is the risk that our results could be more related to the different types of lidars that were used, rather than the geographic effect. We hope to publish soon a test of the method over an area located in another sea or ocean.

42.Figg. 16, 17: the y-axis labels are not specific enough.

ok

43.You can combine Figures 18 and 19 into a single one with two panels. Same for 20-21 and 22-23, and 24-25.

ok

44.A data or code&data availability statement is missing.

The data are available online (SAR image from ESA and Lidar data from RVO). However, due to the corporate constraints we have, unfortunately, we cannot provide our code.

45.A conflict of interest statement is missing

We have no conflict of interest

Best regards, the authors.

22/06/2021