Revised version of manuscript WES-2021-35: answer to comments

Dear Editor,

We finalized the revised version of WES-2021-35 and would like to draw your attention to the following points:

- Regarding Anonymous Referee 1 first major comment (replicability), we gave more details about the algorithms, the data, and the numerical model. We also improved the clarity of the paper and the definition of the statistical criteria. Most of these changes are explained in the answers to Anonymous Referee 1 minor comments related to this first major comment.

- We decided to present only the two-step method (a first algorithm correcting the SAR surface wind speeds and a second algorithm extrapolating these wind speeds to hub height) and renounced to present the method combining these two steps into a single one using on-site Lidar measurements. The first reason is that it makes the paper easier to read and reduces the number of figures, as recommended by both Anonymous Referees. The second reason is that the method combining the two algorithms is unapplicable in practice: typically, only one or two Lidars per site are installed for a short period to assess the resource, which is insufficient to create a training dataset.

- Regarding Anonymous Referee 1 second major comment about the 'a posteriori' correction of the wind speed standard deviation, we could identify the source of the problem. The wind speed standard deviation is underestimated because the machine learning algorithms are designed to provide only expected values. Therefore, they remove some variability related to the errors. In order to correct this effect and maintain the original variability of the wind speed, we artificially added a random variable having the same distribution as the errors. This appears to solve the problem and no 'a posteriori' correction is needed anymore.

Again, we would like to thank you and the Anonymous Referees for your contributions. Please find the detailed answers below.

Best regards,

The authors.

Response to Anonymous Referee 1

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.

We gave more details on the machine learning in the revised version, including the full hyperparametrization. We also gave more details about the data and the numerical model. However, due to ATMOSKY corporate constraints, we cannot provide the full PBL parametrization of the WRF model. We think this is not a limitation since it is just used as source of meteorological parameters related to the atmospheric stability and is not the main purpose of this paper. However, we provided the references that were used to choose the PBL scheme.

Regarding the term 'error', we clearly explained to which difference it refers each time it is used.

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.

We found out that the main reason why the wind speed standard deviation is underestimated. It is because the machine learning algorithm provides the expected value of the wind speed, which removes a part of its variability related to the errors. This does not affect the mean wind speed but squeezes its distribution. In order to compensate for that effect, we added a random variable having the same distribution as the error to the algorithm outputs. For each data point, we created 5 artificial output datapoints by adding different realizations of this random variable mimicking the errors. This helps the wind speed standard deviation to converge towards its definitive value. This method leads to an almost unbiased wind speed standard deviation and no 'a posteriori' correction is needed.

SPECIFIC COMMENTS:

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

L. 32: Currently, it is estimated by using numerical models and/or Doppler wind Lidars installed at the sea surface pointing upwards (NREL, 2020).

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

We are referring to the limited capability of numerical models to resolve small scale phenomena due to their scale truncation. Moreover, the effect of this scale truncation is also unknown because Navier-Stokes equation are chaotic and unsolved. As a consequence, there is no theoretical formula for the error of numerical models and in-situ validations cannot be extrapolated to other locations or periods.

L. 35: Conversely, numerical models provide outputs over the entire area of interest. However, they are not capable of resolving small scale phenomena due to their physics and resolution. As a result, their errors are not precisely known and may vary in time and space. This is particularly problematic in coastal areas where processes are more complex and involve smaller scales.

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

L. 32: Currently, it is estimated by using numerical models and/or Doppler wind Lidars installed at the sea surface pointing upwards (NREL, 2020).

NREL. Best Practices for the Validation of U.S. Offshore Wind Resource Models. Optis, M., Bodini, N., Debnath, M., and Doubrawa, P., 2020, https://www.nrel.gov/docs/fy21osti/78375.pdf, last accessed 24 August 2021.

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

From the point of view of this study, we prefer to call lidars simply 'in-situ' instruments, 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.

L. 56: (in the context of this study, the term 'in-situ instruments' includes profiling Lidars, although technically they use remote sensing)

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

L. 79: Section 2 describes the SAR data used in this study, the numerical model, the Lidar data used as a reference to train the algorithms, and the formulas used to compute the wind power. Section 3 describes the two machine learning algorithms designed to improve the accuracy of SAR surface winds and extrapolate them to hub height, respectively. The reason for separating the method into two algorithms is the scarcity of offshore Lidar data. Since the first algorithm correcting SAR surface wind biases depends on geometric properties of the sensor, it may be improved by using a large network of classical metocean buoys as a training dataset in the future. On the contrary, the algorithm extrapolating surface winds to higher altitudes only depends on meteorological parameters. Therefore, it can be trained with a few Lidars in one location and applied in other areas (if similar meteorological conditions are met). In Section 4, the method is tested in two areas off the Dutch coast where profiling Lidar data are available. The SAR wind speeds extrapolated at hub height are converted into a Weibull distribution, and the extractible power is obtained by simulating the presence of a typical 10 MW wind turbine operating at 200 m. The resulting maps are presented and compared with the output of the numerical model in order to estimate the benefit of using this method compared with a state-of-the-art technique.

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

L. 95: The WRF (Weather Research and Forecasting) non-hydrostatic meso-scale model (Skamarock et al., 2019) was run over these areas with a resolution of 1 km.

We gave the reference mentioned in WRF website 'how to cite' section:

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, Z. Liu, J. Berner, W. Wang, J. G. Powers, M. G. Duda, D. M. Barker, and X.-Y. Huang: A Description of the Advanced Research WRF Version 4. NCAR Tech. Note NCAR/TN-556+STR*,* 145 pp., doi:10.5065/1dfh-6p97, 2019.

7. Please add a reference for GFS, too.

No reference could be found on GFS website. We can only say it was developed by NCEP.

L. 98: It was fueled at its boundary limits by a larger-scale model, the reanalyzed GFS (Global Forecast System) having a resolution of 0.5° developed by NCEP (National Centers for Environmental Prediction).

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.

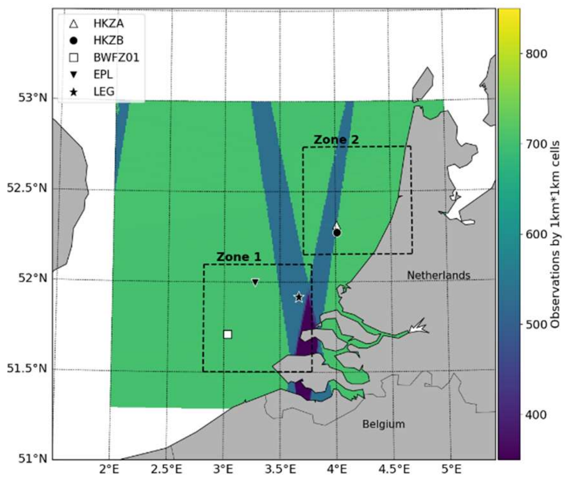


Figure 1: Locations of Zone 1 (bottom, latitude 51.50°N - 52.09°N / longitude 2.82°E - 3.77°E) and Zone 2 (top, latitude 52.15°N - 52.74°N / longitude 3.71°E - 4.68°E) with the positions of the profiling Lidars. The colour represents the number of Sentinel-1 SAR Level 2 wind observations during years 2017, 2018 and 2019.

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 provided the reference that was used to do it.

L. 96: The Planetary Boundary Layer (PBL) parametrization of the model was based on Hahmann etal., 2020.

Hahmann, A. N., Sīle, T., Witha, B., Davis, N. N., Dörenkämper, M., Ezber, Y., García-Bustamante, E., González-Rouco, J. F., Navarro, J., Olsen, B. T., & Söderberg, S.: The making of the New European Wind Atlas – Part 1: Model sensitivity. Geoscientific Model Development, 13(10), 5053-5078, doi: 10.5194/gmd-13-5053-2020, 2020.

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/

We used Viridis colorbar.

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.

L. 116: The data were quality checked by our data provider C2WIND (for each time intervals, the minimum number of packets was set at 20 and the minimum availability at 80%).

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.

L. 111: The dataset used in this study comprises five ground-based profiling Lidars located off the Dutch coast (Figure 1).

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

Lidar	Longitude	Latitude	First date	Last date	Number of levels	Lowest altitude	Highest altitude
HKZA	4.011°E	52.309°N	2016-06-05	2018-06-05	11	30m	200m
HKZB	4.013°E	52.292°N	2016-06-05	2018-06-05	11	30m	200m
LEG	3.667°E	51.917°N	2014-11-17	2017-03-31	10	61m	300m

EPL	3.276°E	51.998°N	2016-05-30	2017-03-31	11	61m	290m
BWFZ01	3.033°E	51.71°N	2015-06-11	2017-02-27	10	30m	200m

Table 1: Main characteristics of the five profiling lidars

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

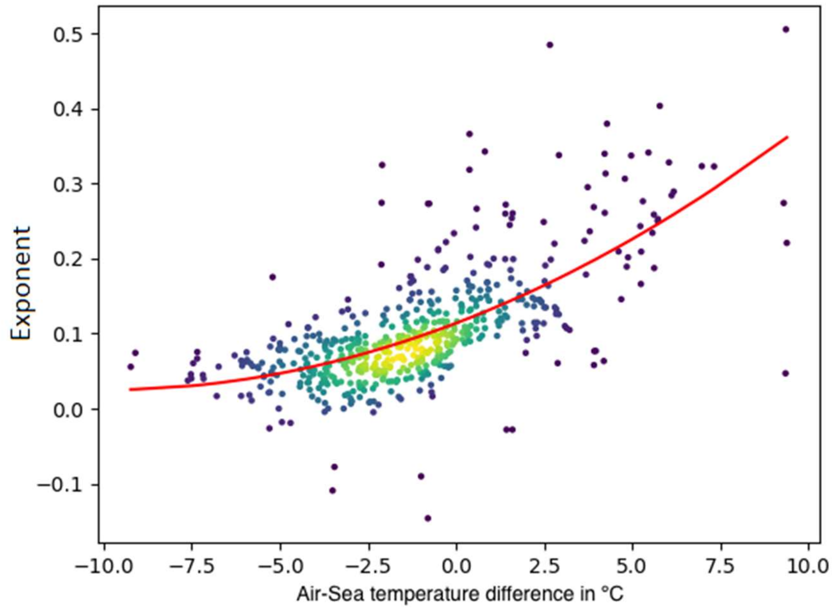


Figure 2: Exponent of the power law between the wind speeds at 4 m and 40 m as a function of the air-sea temperature difference fitted with a second-degree polynomial fit (red curve) with the following coefficients: Y= 0.1137 + 0.0178 X + 0.001 X^2. The colours represent the density of points.

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

Ok. 'Correct on average' means that is the exponent 0.11 is unbiased (the mean error is 0). However, we decided to refine it depending on the atmospheric stability because it can fluctuate importantly between stable and unstable conditions.

L. 130: Hsu et al. (1994) recommend choosing an exponent of 0.11 over the sea. We checked this hypothesis with HKZA and HKZB Lidars that were equipped with anemometers measuring wind speed at 4 m a.s.l.. This exponent was found to be indeed correct on average. However, in order to refine the wind speed values extrapolated at 10 m a.s.l., we adapted the exponent depending on the current atmospheric stability.

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 do not think a histogram would be necessary here.

L. 152: The revisit rate is one passage every two days, which occurs usually in the morning around 5 AM or in the evening around 5 PM (UTC). The satellites pass in the morning or in the evening depending on the orbit orientation, descending or ascending, respectively. The exact acquisition time can vary by plus or minus 30 mn depending on the incidence angle under which the region of interest is observed.

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).

Figure 3 was removed.

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.

Figure 4 and 5 were removed.

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

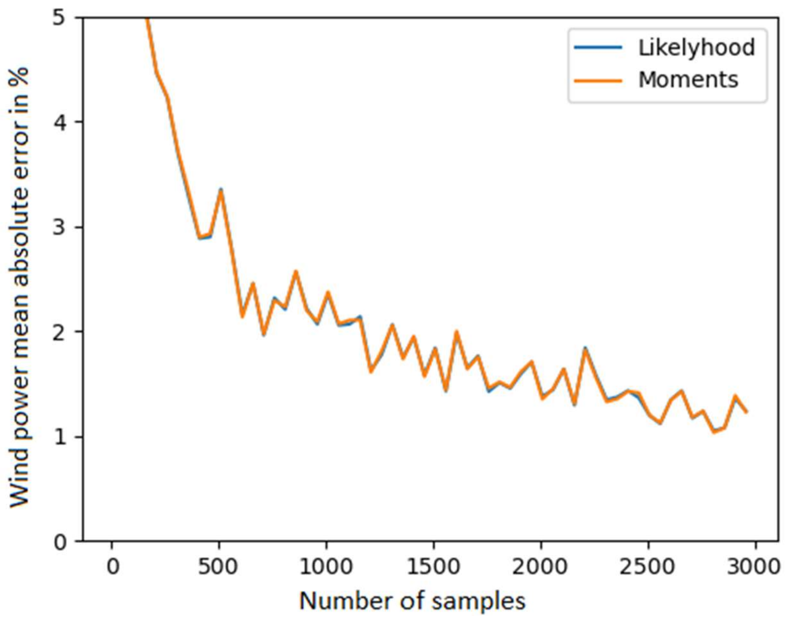


Figure 3: Wind power mean absolute error in percentage as a function of the number of samples, using maximum likelihood to fit the wind Weibull pdf (orange curve), or the method of the moments (blue curve).

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

L. 213: In order to verify this, we simulated the satellites' passages over the Lidars by computing the mean wind speed and the wind power using only the Lidar measurements at the satellites' passage times. These values were compared to those obtained using all Lidar measurements at any time of day. For each Lidar, the differences were found to be below 1% and 2%, respectively, for the mean wind speed and the wind power.

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.

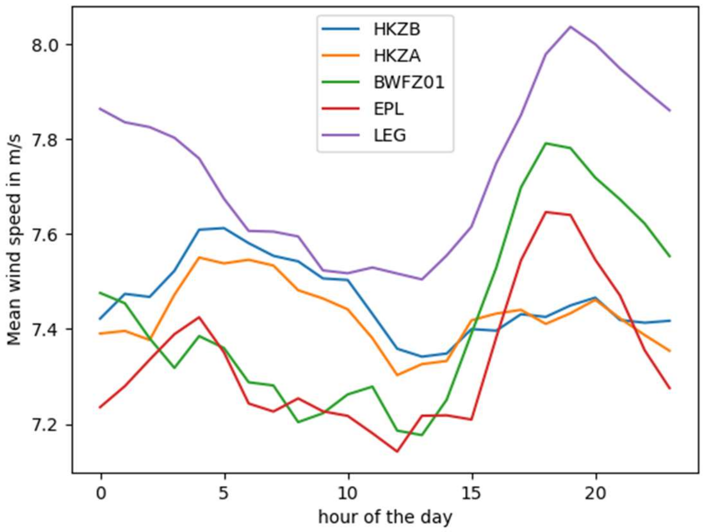


Figure 4: Intra-diurnal variability of the mean wind speed at 10 m for each Lidar. The time is given in UTC, which is close to the local time since Zone 1 and Zone 2 are located near Greenwich meridian.

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.

Regarding global numerical models, there is no debate that in-situ real data are more accurate. It is also obvious that coastal processes involve smaller scales than the ones in open seas, thus increasing potential errors due to their scale truncation. This is why high-resolution models like WRF are usually used to simulate coastal processes.

L. 229: In addition, GMFs were empirically designed using global numerical models as a reference, but they are not as reliable as real data, especially in coastal areas.

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.

L. 236: Given the complex relation between the sea state and the wind speed, and the number of factors able to influence it, machine learning was found to be an appropriate technique to improve the accuracy of SAR surface wind speeds and remove their biases.

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

There seems to be a misunderstanding between us and the Anonymous Referee about the definition of 'errors' and 'bias', maybe due to different versions of English language. For us the errors of the SAR are simply defined as the instantaneous errors compared to Lidar measurements, and the bias is the mean of these errors.

L. 245: In order to select the input parameters, we made a list of interesting parameters and looked for the ones related to the differences between SAR and Lidars. This was done visually by plotting scatterplots of these parameters against the errors of the SAR compared to Lidar measurements.

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 evaluated the parameter importance, so we checked the accuracy of our choices afterwards and a visual inspection is therefore enough here.

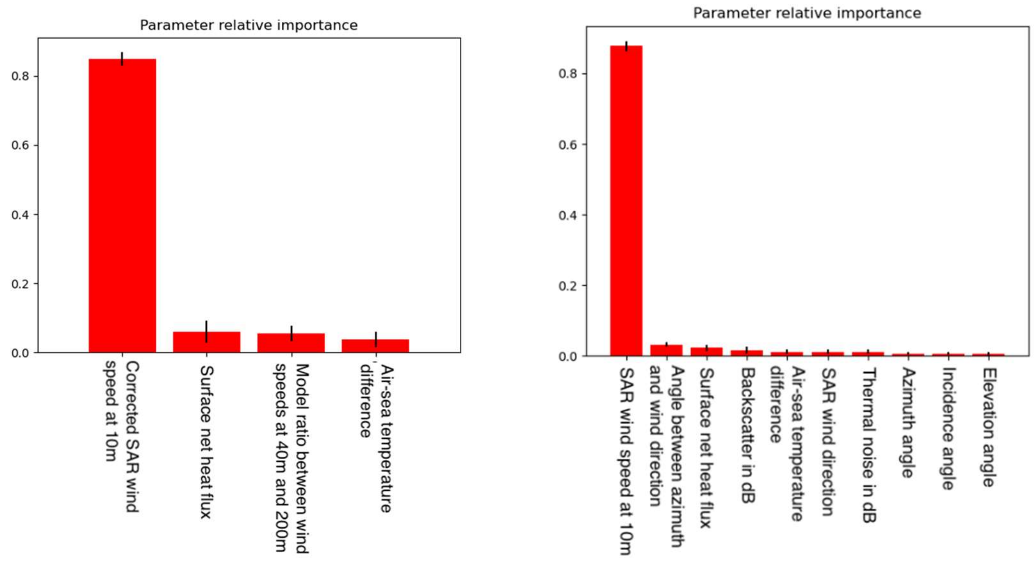
L. 246: This was done visually by plotting scatterplots of these parameters against the errors of the SAR compared to Lidar measurements.

L. 253: The relative importance of these parameters was measured after the training stage using the feature_importances_ attribute of Scikit-learn Python toolbox (Figure 5).

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.

L. 247: The following parameters were selected: the SAR surface wind, the SAR wind direction, the azimuth angle (i.e., the angle between the North and the satellite track), the incidence angle (i.e., the angle between the radar illumination and the zenith of the target), the elevation angle (i.e., the angle between the radar illumination and the nadir of the satellite),…

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…?

L. 241: The algorithm was trained with the wind measured by the Lidars extrapolated to 10 m (the first Lidar level was extrapolated to this altitude with a power law, see Section 2.2). Combining all measurement sites, more than 1000 collocated data points between the Lidars and Sentinel-1 SAR could be found. The algorithm was trained with 50% of the data points randomly chosen, and the rest of them were used as a test dataset.

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?

We dropped neural networks in order to simplify the paper and because it is well known that Random Forest usually performs better than them in regression tasks.

L. 238: We used a Random Forest algorithm (Breiman, 2001), which is known to perform well in regression tasks. It was implemented with the RandomForestRegressor function of Scikit-learn Python toolbox and its architecture was chosen by using cross-validation. The default hyperparameters were found to be the most appropriate ones, except the number of trees set to 240 and the maximum depth set to 20. The algorithm was trained with the wind measured by the Lidars extrapolated to 10 m (the first Lidar level was extrapolated to this altitude with a power law, see Section 2.2). Combining all measurement sites, more than 1000 collocated data points between the Lidars and Sentinel-1 SAR could be found. The algorithm was trained with 50% of the data points randomly chosen, and the rest of them were used as a test dataset.

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). See https://en.wikipedia.org/wiki/Bias_(statistics). For us, 'mean bias' would not be correct.

33.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.

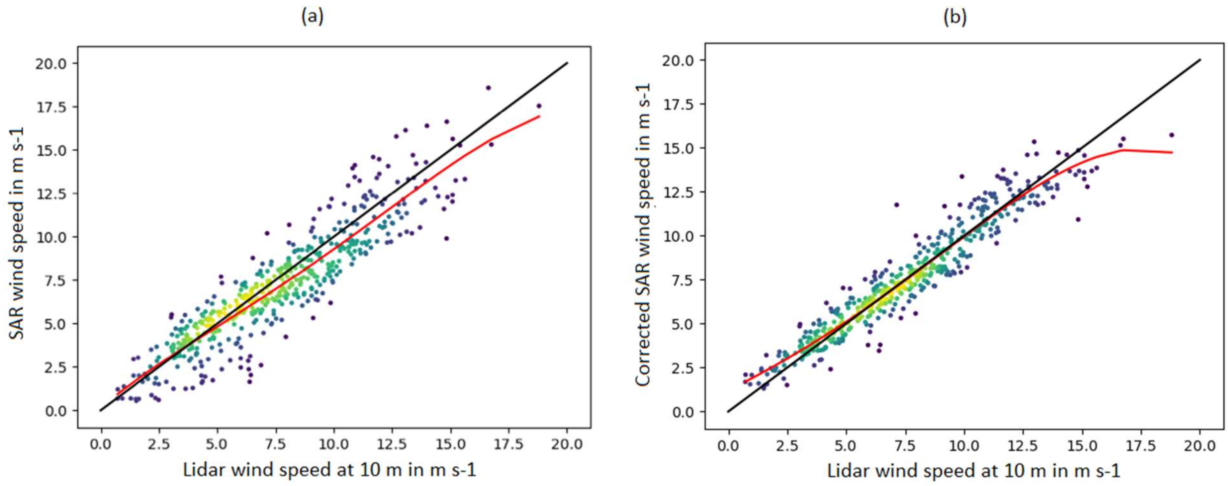


Figure 6: Scatterplots between the SAR and Lidar wind speeds at 10 m before machine learning (a) and after machine learning (b) using the test dataset. The colours represent the density of points. The black curve is the identity line and the red curve a fourth-degree polynomial fit illustrating the bias.

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.

Figure 10 was removed

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

L. 58: Regarding the extrapolation of surface wind speeds to higher altitudes, the statistical theory of turbulence provides theoretical wind profiles (see, e.g., Grachev and Fairall, 1996). However, the problem has not been satisfactorily resolved and becomes increasingly critical as the typical height of windmills increases. Empirical evidence from offshore meteorological masts measurements suggests that a simple power law could be sufficient to model the wind profile (Hsu et al., 1994). Nevertheless, the analysis of Lidar data shows that, above 40 m, the power law is no longer accurate. This limitation has led some authors to use numerical models to improve the extrapolation to higher altitudes (Badger et al., 2016). The advantage of numerical models is that they provide information about atmospheric stability through parameters like surface temperature and surface heat flux. In Badger et al. (2016), these surface parameters were averaged and combined with the similarity theory of Monin-Obukhov to extrapolate wind Weibull parameters. However, to our knowledge, this method was validated with only one meteorological mast in the Baltic Sea and not higher than an altitude of 100 m. Therefore, more research is needed to improve the estimation of wind resources at hub height with SAR data, and convince the industry to use them.

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.

L. 286: The algorithm was also implemented with the RandomForestRegressor function of Scikit-learn Python toolbox. We used the default hyperparameters, except the number of trees set to 340, the maximum depth set to 50 and the maximum number of features set to 'sqrt'. The relative importance of the parameters after the training phase is shown in Figure 7.

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

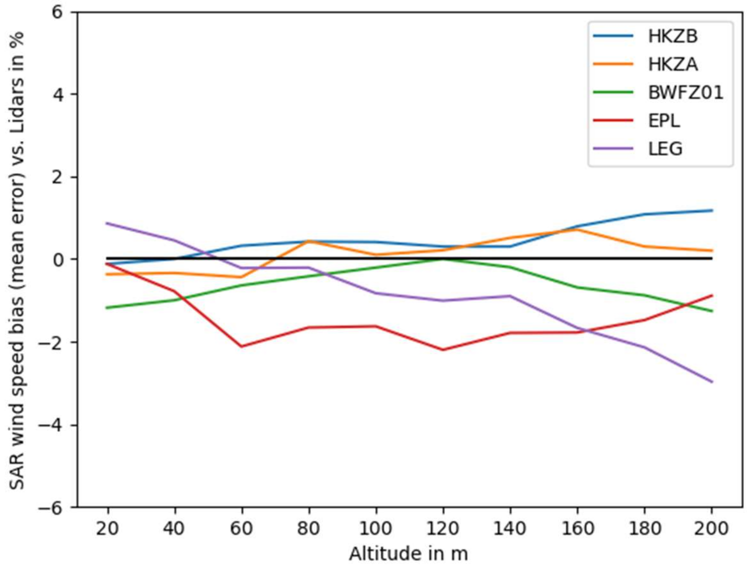


Figure 8: Bias of the extrapolated SAR wind speed against each Lidar in percentage.

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. 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. We do not have a specific explanation on why the bias is positive, negative, going upward or downward depending on the Lidar.

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

Figure 14 was removed (we assume you were 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.

L. 316: In order to compensate for this effect, we reintroduced artificially the original variability of the data. This was done by analysing the distribution of the SAR wind speed errors compared to Lidar measurements and adding a similar random variable to the SAR wind speed obtained after machine learning. The appropriate random variable was found to be a Gaussian with the standard deviation of the SAR wind speed errors. For each data point, at least five additional artificial datapoints needed to be created for the wind speed standard deviation to converge. After this bootstrap, the wind speed standard deviation error was 1.5% when considering all Lidars together in the test dataset. Thus, the result of this correction is an almost unbiased estimation of the wind speed standard deviation.

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 sea 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.

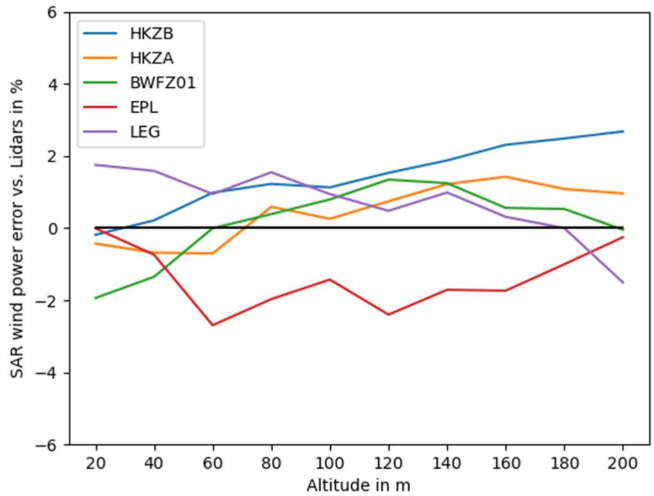


Figure 10: SAR wind power error in % compared to the one computed with Lidars measurements.

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.

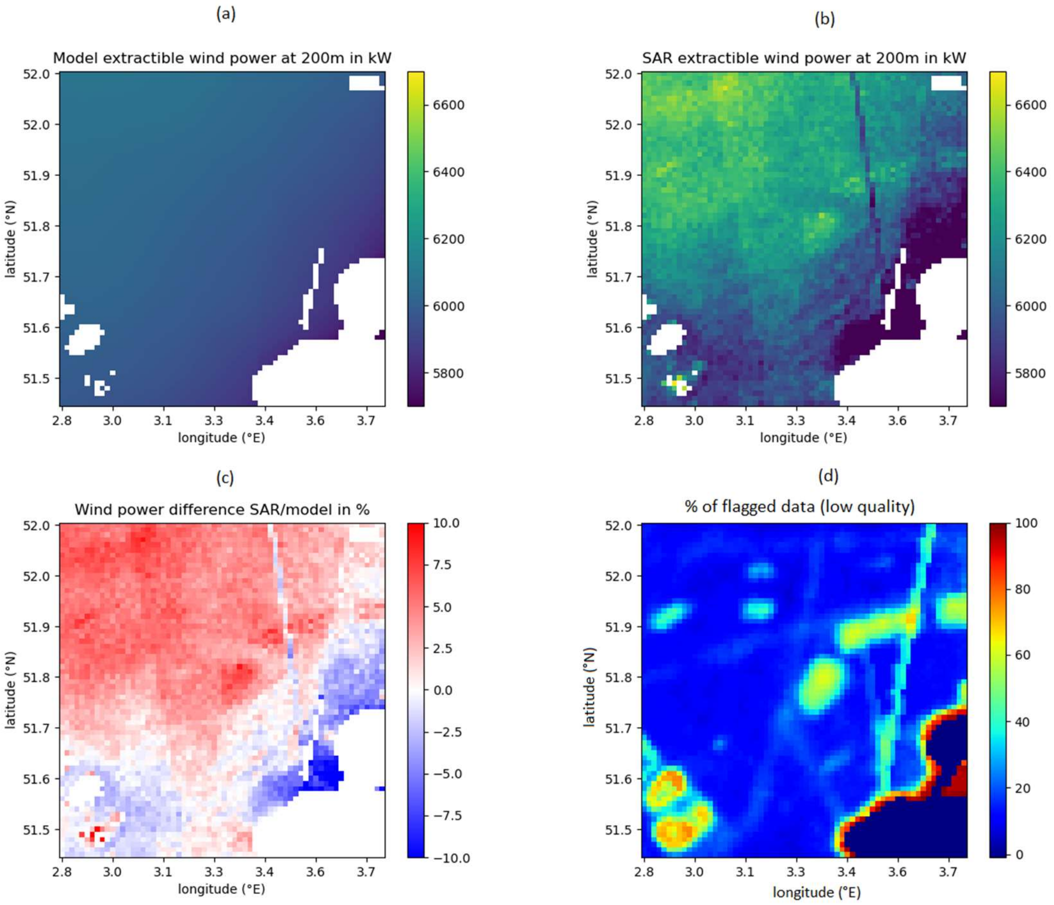


Figure 11: Extractible wind power over Zone 1 in kW for a typical 10 MW turbine predicted by the numerical model (a) and SAR satellites (b), difference in percentage (c), and percentage of low-quality SAR data (d).

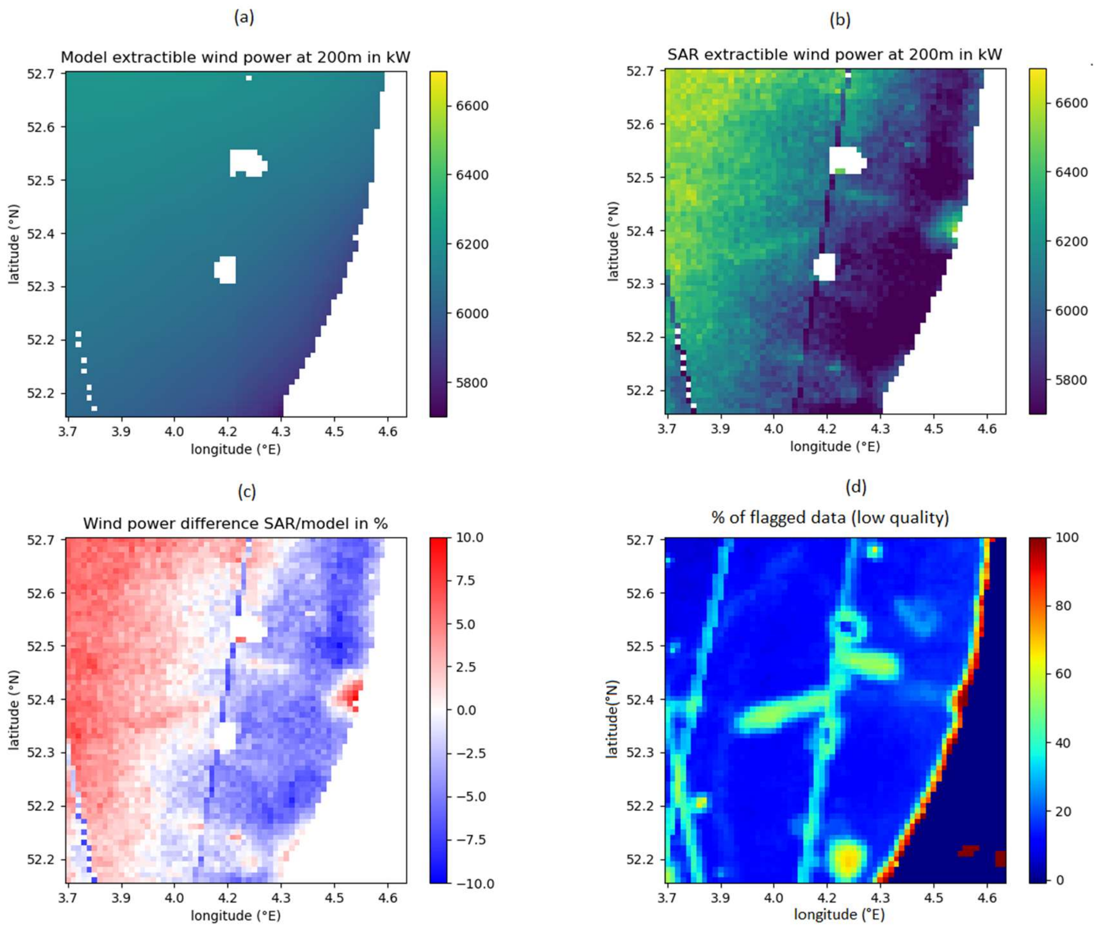


Figure 12: Extractible wind power over Zone 2 in kW for a typical 10 MW turbine predicted by the numerical model (a) and SAR satellites (b), difference in percentage (c), and percentage of low-quality SAR data (d).

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

L. 390: **Author contribution**

Louis de Montera designed the algorithm and wrote the paper, Henrick Berger processed the SAR raw data and created a Level 2 gridded wind product. Romain Husson provided his expertise on SAR satellite and wind measurement from space. Pascal Appelghem parametrized the WRF model and performed the runs. Laurent Guerlou and Mauricio Fragoso supervised the study, organised the funding, and gathered together the project team.

Competing interests

The authors declare that they have no conflict of interest.

Response to Anonymous Referee 2

Review of manuscript "High-resolution offshore wind resource assessment at turbine hub height with Sentinel-1 SAR data machine learning" by Louis de Montera et al.

As I am providing the second review of the manuscript, and have had the change to go through the first reviewer's comments (referred to as RC1), I can start with pointing out that I fully agree with the raised criticism and recommended changes for an improvement of the manuscript.

Below I list my own major and minor comments, some of them being a repetition of those in RC1 but also including some additional input:

(comments in order of appearance in manuscript)

[l 11] When referring to "Lidar measurements", first time here in the abstract, please specify what kind of measurements you mean explicitly – e.g. Doppler wind lidar, somewhat ground-based, measurements of wind velocity profile in range relevant for wind energy applications, or similar.ll Please check the overall manuscript for a sufficiently specific terminology with this respect.

L. 10: The method is tested in two 70 km x 70 km areas off the Dutch coast where measurements from Doppler wind Lidars installed at the sea surface are available and can be used as a reference.

[ll 13-14] When stating a bias, you also need to mention the considered reference – please add this here.

L. 13: The SAR wind speed bias against Lidar measurements at 10 m above sea level is reduced from -0.42 m s^{-1} to 0.02 m s^{-1}, and its standard deviation from 1.41 m s^{-1} to 0.98 m s^{-1}.

[ll 19-20] When reading the sentence "The accuracy of the wind power…" it becomes not clear how you get to the numbers you compare. You should also refer to the process of deriving a power curve from (any) wind data, i.e. the involved derivation.

L. 17: Once the wind speed at turbine hub height is obtained, we assume the presence of an 10 MW turbine with a typical power curve. The extractible wind power is calculated by obtaining the wind speed Weibull distribution with the method of the moments, and then multiplying it by the turbine power curve.

I am also wondering if you really need to consider this (as I understood later, very simplified power curved derivation) for your study, or instead could focus on a derivation of wind power density.

The problem with deriving a total wind power density is that we would have a higher error when comparing with Lidars. When multiplying the Weibull pdf by a power curve, which is what is done in practice, the error is much lower. This is why we chose to use a power curve. More precisely, the error is higher when computing the total wind power density, because one has to consider very low wind speeds and very high winds speeds. In these ranges, wind speed is difficult to measure precisely with SAR satellites. Our opinion is that, in these ranges, the turbine is usually not functioning anyway, so that there is no need to consider them in practice when looking at the extractible power. So, to sum up, we chose this approach because it is closer to industrial applications and because it is more appropriate to test the use of SAR satellites. Note that we now use a more realistic power curve (the one of DTU 10MW reference turbine).

[l 39] I believe you should refer here to ground- (or bottom-) based Lidars.

L. 42: Contrary to ground-based Lidars, spaceborne sensors have the advantage of sounding large areas with high spatial resolution.

[ll 43-44] The sentence seems incomplete – add "… of meteorological conditions [that may impact this extrapolation]", or similar.

L. 46: Moreover, the extrapolation of their measurements from the sea surface to hub height is not an easy task due to the variety of meteorological conditions that may impact the wind speed extrapolation ratio.

[l 59] Be more specific here: "found to give good results" for what explicitly?

L. 70: Regarding their extrapolation at higher altitudes, on land, machine learning has also been found to improve the accuracy of wind speeds extrapolated at turbine hub height compared to power laws or logarithmic laws (Türkan et al., 2016; Mohandes and Rehman, 2018; Vassallo et al., 2019). Optis et al. (2021) also found that machine learning was more efficient at extrapolating offshore winds than theoretical approaches.

[l 89] Here and at other places where you introduce already available models/methods, please add a reference – in this case, for WRF.

L. 95: The WRF (Weather Research and Forecasting) non-hydrostatic meso-scale model (Skamarock et al., 2019) was run over these areas with a resolution of 1 km. The Planetary Boundary Layer (PBL) parametrization of the model was based on Hahmann etal., 2020.

[Figure 1] I suggest to add a larger map to help locating this cut-out. Please also introduce the used abbreviations.

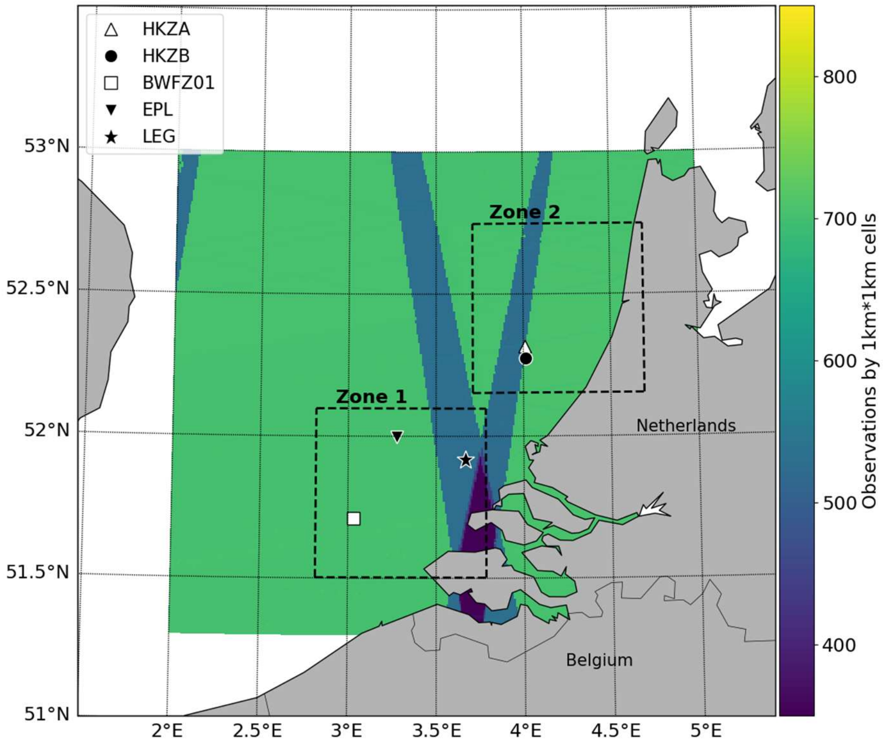


Figure 1: Locations of Zone 1 (bottom, latitude 51.50°N - 52.09°N / longitude 2.82°E - 3.77°E) and Zone 2 (top, latitude 52.15°N - 52.74°N / longitude 3.71°E - 4.68°E) with the positions of the profiling Lidars. The colour represents the number of Sentinel-1 SAR Level 2 wind observations during years 2017, 2018 and 2019.

L. 112: HKZ stand for Hollandse Kust Zuid wind farm, BWF for Borssele Wind Farm Zone, EPL for European Platform, and LEG for Lichteiland Goeree platform.

[Figure 1 and Table 1] Are you sure that all these datasets are from floating lidars? I have at least some doubt with respect to LEG and IJM. Please re-confirm.

After checking, we confirm LEG, EPL and IJM are not floating Lidars but actually on platforms.

L. 114: Lidars HKZA, HKZB, BWFZ01 are floating. Lidars EPL and LEG are installed on platforms.

[l 115] From own experiences and also in line with available recommended practices, I would not fully trust the lower height (as 4 m a.s.l.) measurements from floating lidar systems – mostly from in-situ sensors heavily influenced by the structure itself. Please have some thoughts on this, and possibly consider some added uncertainty.

We decided to use the first level of Lidar profiles and extrapolate it to 10m asl. The power law used to do that which takes into account the atmospheric stability is derived by using the anemometer at the base of the lidars. This might indeed introduce some uncertainty. However, this is still more precise than using a simple power law with a constant exponent. In our case, the uncertainty that is introduced is compensated by the second machine learning algorithm doing the extrapolation. Since the results we provide are given after applying this second algorithm trained with the Lidar measurements at 200m, there is no need to add an additional uncertainty. In the future, we will be using the NDBC buoy network to perform the correction of SAR surface wind speeds, which will avoid this problem and generalize the method to other location.

L. 136: The anemometers located at the base of the Lidars at 4 m a.s.l. do not have a high precision and may add some uncertainty, however, since the final machine learning algorithm presented in this study is trained with Lidar measurements at hub height, this uncertainty is included in our results.

[Figure 3] This figure is not very informative – please review the design and information included. Overall, I think you can and should reduce the number of figures in the manuscript, possibly combining some of them.

Figure 3, 4, 5 were removed

[Figure 4 and Figure 5] In these plots it may be helpful to have the coastline (or something else) as reference.

Figure 3, 4, 5 were removed

[section 2.4] As pointed out above, I think, the used power curve is too simplified. I would suggest to either apply a more realistic power curve, or instead consider another quantity as e.g. wind power density for this investigation.

We used a more realistic power curve:

https://nrel.github.io/turbine-models/DTU_10MW_178_RWT_v1.html

As explained above, we prefer not to compute the wind power density instead since the SAR is less precise in measuring low and high wind speeds.

L. 176: We chose to simulate an 10MW turbine with a typical power curve: the DTU 10 MW Reference Wind Turbine V1 (see DTU Wind Energy, 2017, and https://github.com/NREL/turbine-models/blob/master/Offshore/DTU_10MW_178_RWT_v1.csv, last accessed September 2, 2021).

[Figure 6] I do not think that this figure is really needed, instead you could add more details in the text.

Actually, this figure is needed because it shows that the error of the method is highly related to the number of SAR samples. Since the number of samples is growing with time, and since it is possible to use also other SAR satellites (like ENVISAT or RADARSAT2), this figure show that it is possible to improve the accuracy of this method.

[section 2.5] I am confused by your mentioning of “one passage every two days” and “passage times are separated by 12 h” – please be more specific here.

Yes, this is confusing. The satellite passes every two days, and this can occur at specific times: or in the morning or in the evening. The two possible time have a difference of 12h.

L. 152: The revisit rate is one passage every two days, which occurs usually in the morning around 5 AM or in the evening around 5 PM (UTC). The satellites pass in the morning or in the evening depending on the orbit orientation, descending or ascending, respectively. The exact acquisition time can vary by plus or minus 30 mn depending on the incidence angle under which the region of interest is observed.

L. 209: Therefore, since the satellites pass at two possible times of the day separated by 12 h, according to the Nyquist-Shannon sampling theorem, they should be able to capture the majority of the intra-day variability.

[l 222] As already stated above, please add reference for the applied methods – here the “two types of machine learning regressor[s]”.

L. 237: We used a Random Forest algorithm (Breiman, 2001), which is known to perform well in regression tasks. It was implemented with the RandomForestRegressor function of Scikit-learn Python toolbox and its architecture was chosen by using cross-validation.

[l 229] Please specify how “the relative importance” is defined and derived.

The relative importance is based on the mean decrease in impurity (over the trees).

L. 253: The relative importance of these parameters was measured after the training stage using the feature_importances_ attribute of Scikit-learn Python toolbox (Figure 5).

[l 239] Also the statement “Random forest was found to outperform neural networks” needs more explanation / details.

The reference to neural networks was dropped since it does not bring any interesting insight to the reader. It is well known that Random Forest usually outperforms them for regressions tasks.

[Figure 9] Add a legend to the plots and the details of the red curve (fitting parameters).

We did not add that fitting parameters on Figure 9 (now 6) since the fit is just there to illustrate the bias. However, we gave the fitting parameters of Figure 2 since this fit is used to extrapolate wind speeds to other altitudes.

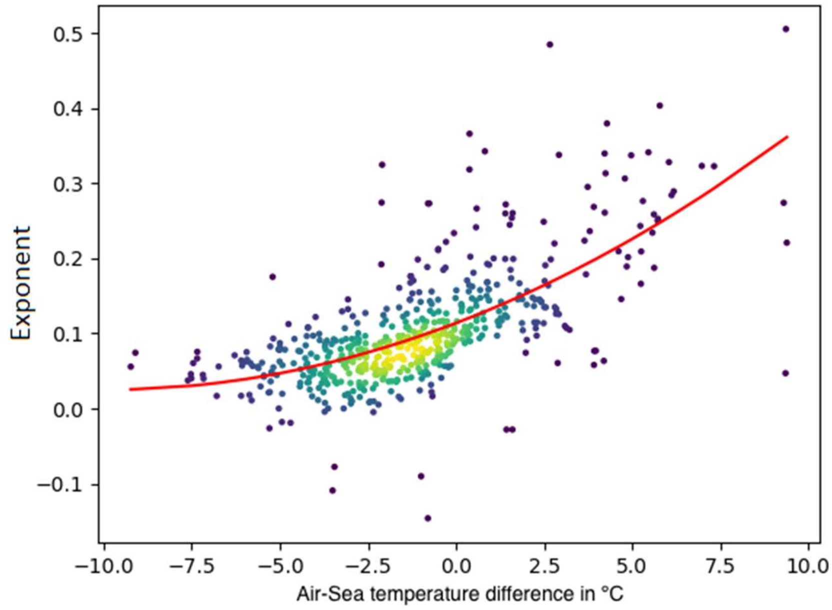


Figure 2: Exponent of the power law between the wind speeds at 4 m and 40 m as a function of the air-sea temperature difference fitted with a second-degree polynomial fit (red curve) with the following coefficients: $Y = 0.1137 + 0.0178\,X + 0.001\,X^2$. The colours represent the density of points.

[l 266] Again, “machine learning” needs more explanation and details.

L. 283: These parameters were used together with the corrected SAR wind speeds at 10 m as input to the Random Forest algorithm, which was trained to learn the Lidar wind speed at several altitude levels until 200 m using the same training dataset as previously. The algorithm was also implemented with the RandomForestRegressor function of Scikit-learn Python toolbox. We used the default hyperparameters, except the number of trees set to 340, the maximum depth set to 50 and the maximum number of features set to ‘sqrt’. The relative importance of the parameters after the training phase is shown in Figure 7.

[Figure 12 (and Figure 14)] It is not clear to me, why you have not used a smaller range for the vertical axis – please revise.

We used a -6% to +6% axis.

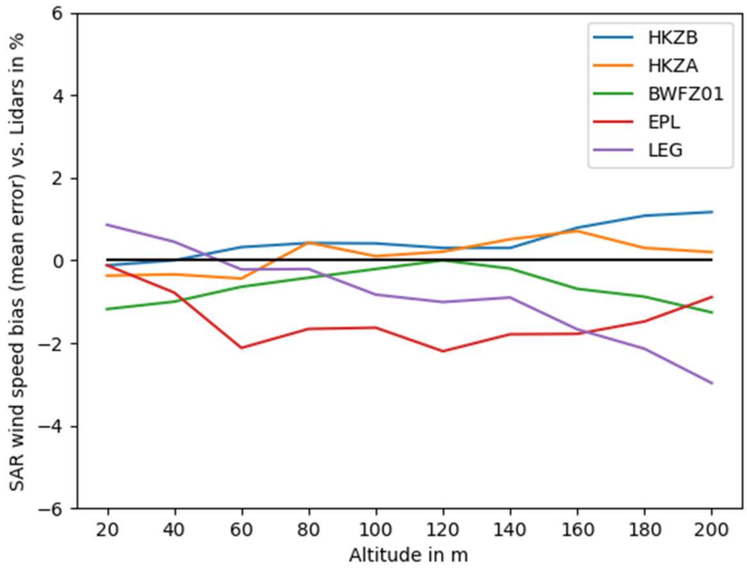


Figure 8: Bias of the extrapolated SAR wind speed against each Lidar in percentage.

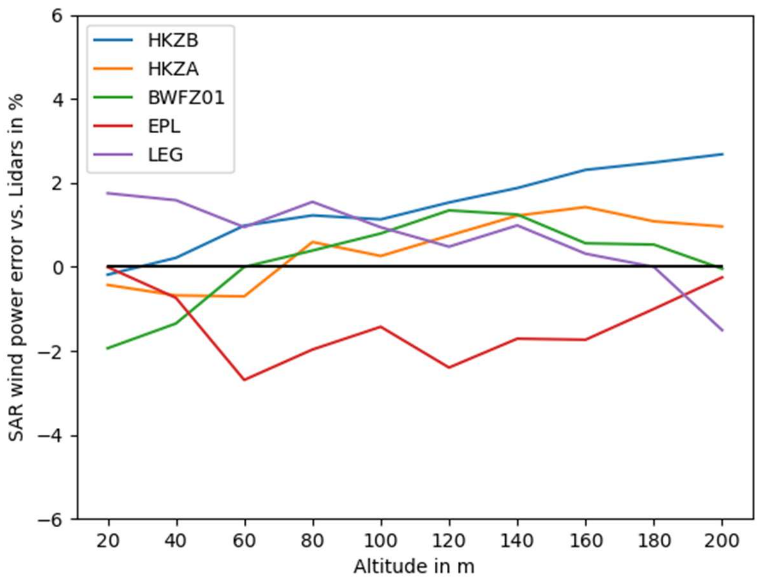


Figure 10: SAR wind power error in % compared to the one computed with Lidars measurements.

[l 323] Sentence "In this case, …" is incomplete.

The word 'the distribution' was missing.

L. 323: In this case, the distribution was shifted to the left, which means that the numerical model underestimates the wind speed compared to Lidar measurements.

[Figure 18 and following] Please re-arrange these plots for better comparability – combine several plots in one figure, for instance.

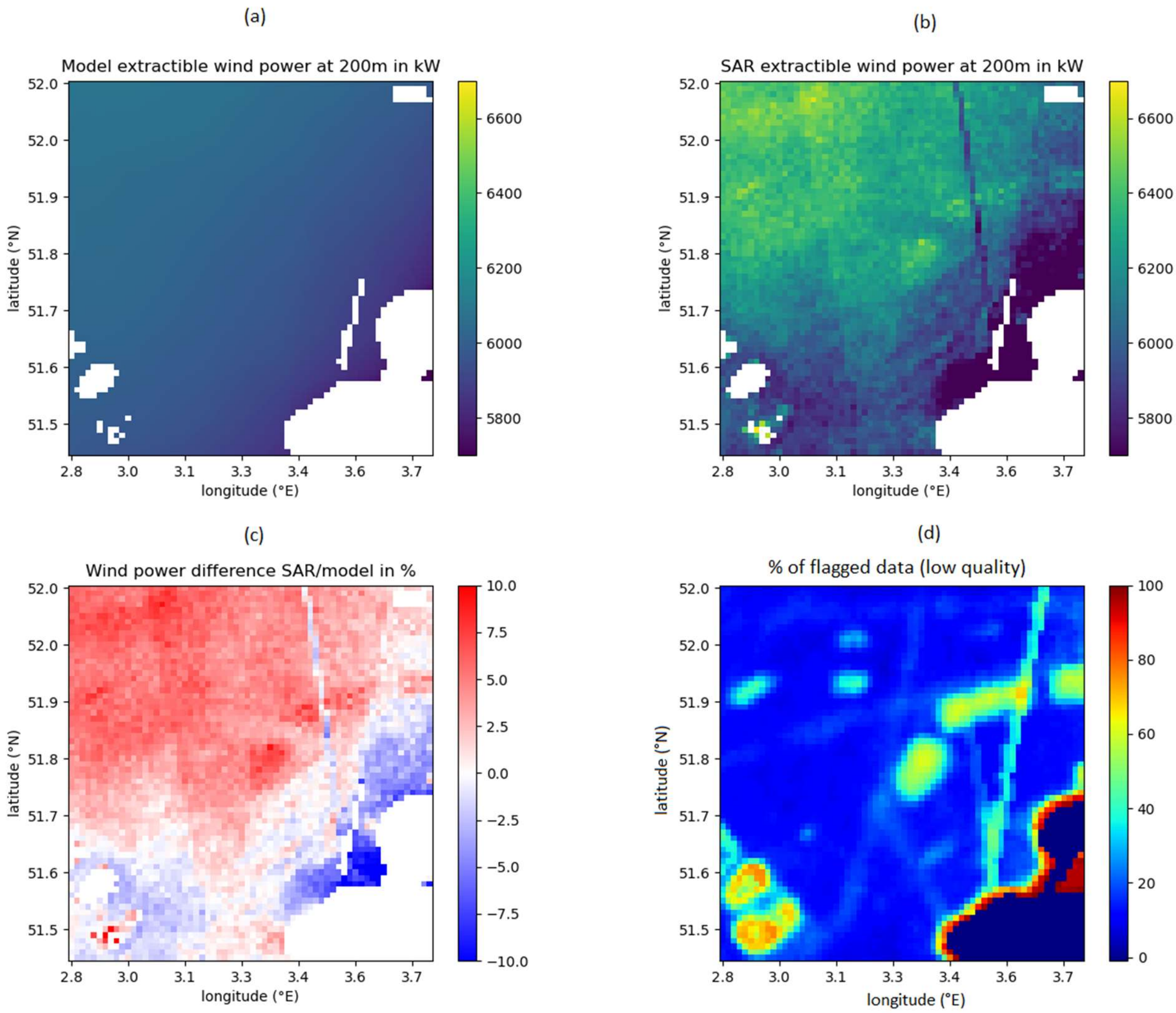


Figure 11: Extractible wind power over Zone 1 in kW for a typical 10 MW turbine predicted by the numerical model (a) and SAR satellites (b), difference in percentage (c), and percentage of low-quality SAR data (d).

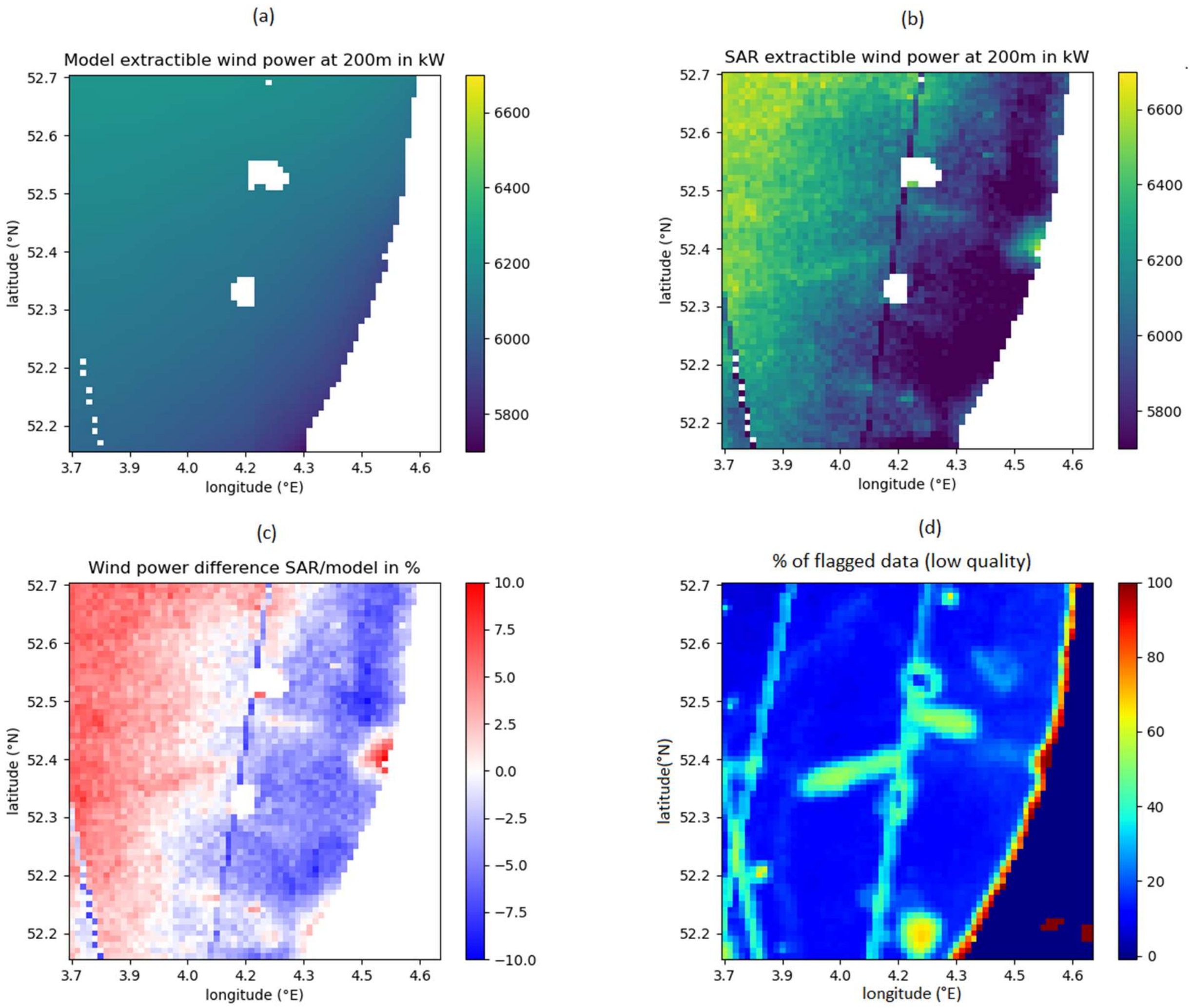


Figure 12: Extractible wind power over Zone 2 in kW for a typical 10 MW turbine predicted by the numerical model (a) and SAR satellites (b), difference in percentage (c), and percentage of low-quality SAR data (d).