Revised version of manuscript WES-2021-35 and answers to comments

Editor's letter:

Dear Authors,

The two reviewers raised several critical issues, and one opted for "reject". Hence, it is clear that the revised manuscript needs a new substantial revision to become publishable. The most crucial issue is your proposed method's applicability and scientific soundness. In addition, it seems you have not satisfactorily responded to all the comments of the first revision round.

I think the paper still has some valuable aspects. So, I suggest that you submit a detailed plan of the changes you propose to make and respond to all the issues from the previous and past round of comments, and we will continue with the review process. If you cannot comply with the requested changes, you can withdraw your submission.

Best regards,
Andrea Hahmann

Response:

Dear Editor,

First of all, we would like to thank you for accepting the management of this review process.

Concerning the paper itself, we would like to say that we did answer all the comments from the first revision round, and especially the major ones. You can have a look to the documents to confirm this. The so-called 'critical flaws' pointed out by Referee1 are actually new major comments from him or her. Therefore, Referee1 advice to reject the paper is not related to a supposed lack of response regarding the first round of review.

Actually, our paper is a major breakthrough since it is the first time SAR can be used to estimate wind resource at hub height with a sufficient accuracy to convince the industry. Moreover, the paper presents a stand-alone method to correct SAR surface winds and another stand-alone method to extrapolate surface winds to higher altitudes. Therefore, we definitely want this paper to be published and we propose you a revised version of the manuscript.

Since Referee1 seems not to understand that the first version of the paper was a proof of concept, and encourages us to disclose our operational method, we included several important improvements in the revised version, which are explicitly described in the paper and listed below. Please also consider that we were originally aiming at proposing a second paper describing an operational version of this first proof-of-concept methodology. Given the time we had between the first and the second revision, we now propose to directly describe the operational methodology we have reach, thereby also addressing the major comment from Referee 1.

Additional improvements with respect to previous version:

- The SAR data are now produced with CMOD7 GMF instead of CMOD5n GMF

- The correction of SAR surface wind is done with a network of buoys located in the US

- The correction now uses also as input the cross-polarization backscatter (it improves strong winds retrieval)

- The correction now uses also as input the ECMWF wind speed provided as SAR metadata (it improves low winds retrieval)

- The WRF is now forced with ERA5 1h instead of GFS 3h (it is more adapted to the small areas of study)

- We now use a Gradient Boosting algorithm instead of Random Forest (it is more efficient and does not disrupt the wind speed distribution)

- The validation of the extrapolation is done with a round-robin technique (so that all samples for each Lidar can be used to estimate wind power)

- The relative importance of parameters is computed with ShAP method.

- The error due to SAR low and irregular sampling is corrected exactly and automatically by using the WRF to simulate the satellites' passages

We would like to emphasise that we reject Referee1 statement that our paper cannot be published because our method is not operational. Firstly, the immediate applicability of scientific results has never been a requirement to publish in scientific journals. Secondly, the algorithm now presented in the revised version is fully operational and was recently used by the French Government to assess AO4 and AO5 offshore sites respectively located in Normandy and Southern Brittany.

We believe that the review from Referee1 is unfair and may raise the question of his or her partiality. For instance, we have found that a team that we recommended as referee for our manuscript was planning to do exactly the same study as ours (see the numerous references about future plans of using random forest to extrapolate SAR winds in Optis et al.: New methods to improve the vertical extrapolation of near-surface offshore wind speeds, Wind Energ. Sci. Discuss. 2021 https://wes.copernicus.org/articles/6/935/2021/). Therefore, we suspect that Refeee1 may have a personal interest in preventing or delaying the publication of our results. Due to this context and if this was the case, the reviewing process may require a conflict-free reviewer, or at least some careful discernment from the editor with respect to the referee's review.

Finally, we would like to thank you again for your time and expertise.

Best regards,

The authors

Referee1's comments:

MAJOR COMMENTS:

1. I still cannot see any practical application for the approach the authors proposed in the paper. In the first part of the approach (described in Section 3.1), you 1) extrapolate lidar winds down to 10 m and 2) apply a machine learning model where the SAR-derived wind speed is an input, and the target variable is the 10-m lidar wind speeds. You then use this SAR-corrected winds for the extrapolation to hub-height. Now, to me this approach has two fatal flows. The first fatal flaw I am seeing is that if one needs to have lidar data available (you use them to correct the SAR winds, Section 3.1), why would one need to extrapolate SAR data in the first place, since the lidar provides hub-height winds already? To respond to this concern, I could see an application of this approach when someone only has let's say a 10-m sonic anemometer (whose wind speed observations are used to correct the 10-m SAR winds), without any hub-height observations. But for this application to be possible, the authors would need to test the generalization of the approach they propose with a round-robin validation. In other words, the authors would need to answer the following question: "how accurate is this whole approach when applied at a site (i.e., in my example, where I only have near-surface wind speed observations) different from the one where it has been trained (i.e., in my example, where I have lidar observations which already give me hub-height data)?". Since the authors have multiple observational locations, they could do this exercise, but currently this validation is not done in the paper. However, even if this validation exercise were to be completed, the second fatal flow I am seeing here would still kick in. If one needs to have any 10-m observation of wind speed to correct for the SAR data, why would one use the SAR data in the first place, instead of just extrapolating to hub-height the 10-m observations coming from the instruments needed to correct the SAR data?

> The first version of the paper was a proof of concept, not yet as an operational product, and we explicitly mentioned in it that we planned to use a buoy network in the future. So we do not see it as a 'critical flaw', but rather as a logical step of our research and development. In any case, since the last round of review the algorithm became fully operational, so we take advantage of this revised version to introduce a correction of SAR surface winds with the NDBC buoy network.
>
> For what concerns what you call the "second fatal flow", we recall that the purpose of the in situ calibration at the sea surface is to be able to correct for systematic errors that are not site specific. They aim at correcting errors related to the GMF and to the sensor calibration. They can therefore be learnt from measurements located elsewhere and at different periods. There is no need to have this source at the location and period of interest, which widens the potential for using it in various locations.
>
> Regarding the extrapolation, we now do a round-robin validation for the sake of increasing the number of sample per Lidar used in wind power estimation. As already explained in the first round of review (in our answer to minor comments), we do not think that the round-robin method adds a lot of value here in terms of validation because the meteorological conditions of the samples are already not correlated (48h minimum time difference) and because the Lidars are too close form each other. In any case, it confirms directly that the method can be trained in one place and applied in another.

SPECIFIC COMMENTS:

1. L. 33: other data sources can be used to estimate hub-height winds, for example reanalysis products.

That's what we mean by 'numerical models'.

2. L. 62: "Nevertheless, the analysis of Lidar data shows that, above 40 m, the power law is no longer accurate." is a strong and very general statement. While it has some merit, it needs references.

We cite in the next paragraph a lot of literature trying to find a more accurate model to extrapolate to hub height. We also checked this with our own Lidar data and confirm the power law is not suitable after 40m. We added a reference (Tieo et al., 2020)

L. 69: 'Nevertheless, above a few tens of meters, the power law model is no longer accurate (see, e.g., Tieo et al., 2020). This limitation has led some authors to use numerical models outputs to improve the extrapolation to higher altitudes (Badger et al., 2016).'

3. Section 2.1 have multiple instances of weird spacing between words.

Ok. Corrected.

4. L. 135: why did you consider two lidars only to determine the exponents of the power law, which are then applied to all the lidars?

It was explained in the previous version of the paper that only these two Lidars had an anemometer at their basis. Anyway, this is not needed in the revised version since we now use the NDBC network and do not extrapolate the Lidar data to a lower altitude.

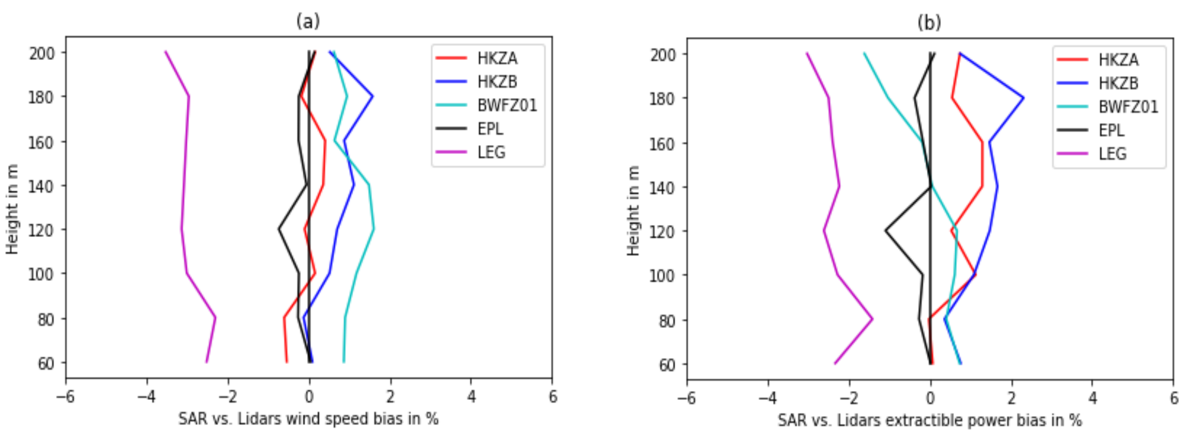
5. L.240: "The default hyperparameters were found to be the most appropriate ones". How did you find this? Remember, your work should be replicable! A similar comment applies to line 286.

We always used Gridsearch (with cross-validation). We mention now it explicitly in the revised version.

L.310: 'The Gradient Boosting hyper-parameters optimized with grid-search are shown in Table 3 (left column). The other hyper-parameters are the default ones. The relative importance of the input parameters is given in Figure 4.'

6. Figure 8: the y-axis label can simply be "SAR – lidar wind speed bias (%)"

Ok



7. Despite my previous comment, a data or code & data availability statement is still missing. You should add one even if your code cannot be shared – simply state it

It is not clear if this section is mandatory in WES guidelines, so we simply added the data providers in the acknowledgment section in the first version. We now added this statement accordingly to your comment. Unfortunately, we are still not allowed to share the code since our company is commercial.

Code and data availability

SAR data are available at ESA. Buoys data are available at NDBC. Lidar data are available at the Dutch Ministry of Economic Affairs and Climate Policy. The WRF source code and Python packages are open source. Unfortunately, the full code of the method developed in this paper is not available due to corporate constraints.

Referee2's comments:

This manuscript is about vertical extrapolation of wind fields from satellite SAR. It is novel and interesting in the context of offshore wind energy projects and planning.

GENERAL COMMENTS:

1. The work is very focused on wind resource assessment and on mapping the wind power potential at the height 200 m. I think, however, that the real advantage of extrapolating instantaneous SAR wind fields lies in the possibility to map the detailed variation of instantaneous winds and compare these with e.g. numerical modeling. This aspect is not mentioned.

> Yes, we need to mention that since it is a really interesting application. Actually, there are 3 important possible applications in the paper: correction of SAR surface winds to create better products, instantaneous extrapolation as you mention, and resource assessment. It is clear from our observations that instantaneous SAR images at hub height also offer additional interesting insights.
>
> *L.25 The algorithms presented in this study are independent from each others and can therefore also be used in a more general context to correct SAR surface winds, extrapolate surface winds to higher altitudes, or produce instantaneous SAR wind fields at hub height.*
>
> *L. 415 The resulting SAR wind speed bias is 0.02 m s^{-1}. Its MAE is 0.57 m s^{-1} and its standard deviation 0.74 m s^{-1}. This algorithm can be used as a standalone to create more accurate SAR wind products. The second algorithm extrapolating surface winds to higher altitudes has been tested against Lidar measurements […] This algorithm can also be used as a standalone to extrapolate wind speeds measured at 4 m above sea level. These two algorithms combined together produce instantaneous SAR wind fields at hub height, which can provide interesting insights to wind farm developers.*

2. A second advantage of the presented approach is that it overcomes the current shortcomings of analytical approaches such as MOST. These are only valid within the surface layer of the atmosphere whereas the machine learning approach can be used at any height as long as there is sufficient data available for training and testing. This aspect is not mentioned.

> Yes, it is clearly another important advantage of machine learning. We modified the introduction accordingly to stress this point.
>
> *L. 84 Moreover, machine learning can be used at any altitude contrary to theoretical approaches that are limited to the boundary layer.*

3. Like previous reviewers, I find that the use of a turbine specific power curve makes the analysis overly complicated and more difficult to follow. I would recommend to map the wind power density instead. If this change is not feasible at this stage, please add some reflections over the effects of using the power curve.

We already answer Referee2 about this point. SAR sensors can have difficulties in detecting very high wind speed and since these wind speeds have a very strong weight in the total power density, SAR estimations of the total power density would be less accurate. So, in order to show that SAR can be used for resource assessment, we chose to use the extractible power instead, which is more accurate since the power curve has a plateau at high wind speeds and is less affected by potential inaccuracies at strongest wind regimes. We consider using the extractible power is not a flaw since the industry computes it from the Weibull parameter in practical applications. We explain it in more details in the revised paper.

L. 232 Since the total wind power density is related to the cube of wind speed, very high wind speeds have a strong influence on its estimation. Since SAR sensors do not detect well very high wind speeds because they tend to saturate, we do not recommend using them to estimate the total wind power density. However, estimating the extractible wind power instead removes this limitation, because wind turbines usually do not operate or function at a plateau when very high wind speeds occur.

4. I think there is some confusion about the term 'hub height'. The 10 MW reference turbine used here has a hub height of 119 m but throughout the manuscript, the 200 m height is described as the 'hub height'.

Ok. We changed all the results to provide maps at 120m. As explained above, the idea was to use a standard power curve in order to remove low and very high wind speeds from the wind power assessment. So we were not focused on the real hub height of the turbine, but just on the shape of the power curve. Therefore, we chose the DTU 10MW turbine, and assumed that a turbine operating at 200m would have more or less the same power curve shape (multiplied by a constant). To avoid such assumption, we produced new maps at 120m.

L. 327 Figures 10 and 11 show the extractible wind power maps at 120 m produced by the WRF and SAR methods assuming a typical 10 MW turbine, and the difference between them in percentage.

5. The quality of the different data sets is not really considered. In particular, I would like to know more about the parameters from WRF: is the accuracy of the instantaneous temperatures and heat fluxes sufficient for this type of analysis. See for instance: Pena Diaz, A., & Hahmann, A. N. (2012). Atmospheric stability and turbulence fluxes at Horns Rev—an intercomparison of sonic, bulk and WRF model data. Wind Energy, 15(5), 717–731. https://doi.org/10.1002/we.500.

Actually, if the numerical were perfect, there would be no need for SAR and machine learning. So we totally agree that the outputs of the numerical model have a low quality. However, this is not the main issue, because the idea here is precisely to learn the errors of the numerical model through machine learning. So our approach was to remove the most unreliable model parameters (like the wind speed), and keep more reliable ones (like the relative extrapolation ratio) or the less fluctuating ones (like heat flux and temperature). By doing so, it seems that we were indeed able to learn the numerical model errors and exploit the remaining information they contain. We explained it better and cited the above-mentioned paper.

L. 207 Since the accuracy of numerical models outputs is questionable, one must be careful when choosing these meteorological parameters. In particular, the WRF wind speed at hub height could not be used directly since the aim of this algorithm is to estimate it with SAR satellites. Instead, we provided the algorithm with the WRF extrapolation ratio between the wind speed at the sea surface and hub height. Using this relative quantity has the advantage of preventing the WRF from interfering with SAR estimates. Moreover, this extrapolation ratio was found to be accurate: the comparison with experimental data shows that its bias was less than 1% for each Lidar. The other relevant parameters related to the atmospheric stability we used were the air-sea temperature difference and the surface heat flux. The accuracy of these parameters is also problematic (see, for example, Pena Diaz & Hahmann, 2012). However, in the context of machine learning, the focus is more on the information they contain, rather than their absolute accuracy. Since they are not fluctuating as quickly as the wind speed, we assumed that their biases were following repetitive patterns that could be learnt by the algorithm, and that these biases would not prevent it from extracting the relevant information.

6. Finally, the presentation of results seems a bit unstructured as results are spread across sections 2, 3, and 4. Sentences alternate between present and past tense. Please be consistent.

Ok, we also created a new section called 'Methods' to clearly separate the methods, data and the results. We will use only past tense.

SPECIFIC COMMENTS:

1. 'it' refers to the error? Perhaps better to state that. (L. 22)

Corrected

L. 17 Once the wind speeds at hub height are obtained, we assume the presence of a 10 MW turbine and estimate the wind Weibull parameters taking into account the SAR irregular temporal sampling. The wind speed Weibull distribution is then multiplied point-by-point by the turbine power curve to obtain the extractible wind power with a 1 km spatial resolution.

2. This description does not fit in here. Perhaps more suitable for Section 2. (L. 81)

Corrected. It was moved to section 3 Methods.

L. 179 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 winds. Since the error depends on the geometry of the sensor, this algorithm was to be trained with a large database of measurements covering the diversity of possible angles obtained from the NDBC network of metocean buoys (Section 2.4).

L. 201 After this correction, the extrapolation of SAR surface winds did not depend on the sensor geometry, therefore, the algorithm could be trained with a dataset including a limited number of instruments, like the Lidar data from the North Sea (Section 2.5).

3. This really depends on a project's level of maturity. A first screening of sites might be based on numerical modeling alone but as a project gets closer to a financing decision, in situ observations are always used, as far as I know. (L. 99)

> Yes, actually we wanted to say that the numerical model was typical of the ones used by the industry, not the whole assessment. In any case, we were already using Lidar data to correct the WRF like the industry is doing (see the previous version of the paper in the results section). In the revised version, we explain now clearly this use of in-situ instruments to correct the WRF bias in the Data section.
>
> *L.141 Moreover, since the WRF is typical of numerical models currently used by industry, we also used it as a reference to assess the benefits of using SAR data (Section 4). Since numerical models are often combined with in-situ measurements to increase their accuracy, we also corrected the WRF bias. The extractible power estimated by the WRF was found to be underestimated by 3% compared to Lidars.*

4. Normally, u and z are used for instantaneous observations (instead of U and Z). (L. 126)

> Corrected

5. I suggest to put the description of SAR data first - before the model and reference data sets. The SAR data represents the core of this work. (L. 146)

> Corrected. It is more logical indeed.

6. I do not think that readers of WES will know the difference between grid spacing and spatial resolution. Please explain or give the grid spacing alone. (L. 148)

> Corrected
>
> *L. 109 Sentinel-1 Level 1 Ground Range Detected (GRD) backscatter product has a spatial resolution of a few tens of meters, whereas Level 2 wind products typically have a spatial resolution of 1 km.*

7. I am a bit confused here: Is the reference used to produced Figure 3 based on a direct calculation of the power from the data set itself without any curve fitting)? It should be. Please add this information. (L. 179)

> The reference used to produce this Figure does not involve any curve fitting. To obtain the extractible power reference, we use the arbitrarily chosen Weibull parameters and the exact formula to get the Weibull pdf. Then we multiply it point-by-point by the 10MW turbine power curve. We explained it with more details and clarity.
>
> *L. 260 The accuracy of this estimation method was assessed with simulations by generating time-series of a Weibull random variable with arbitrary parameters, and then trying to recover the original parameters from these time-series. More specifically, we chose Weibull parameters typical of the North Sea wind climate (k = 2.2 and λ = 8.5) and computed the*

reference extractible power using these parameters and the exact formula (Eq. (2) multiplied point-by-point by the 10 MW turbine power curve). Then, we generated random synthetic wind speed time-series using the Weibull pdf (Eq. (2)) and applied the method of the moment (Eqs. (3) and (4)) to recover the original Weibull parameters and estimated again the extractible power.

8. Again, please specify what is meant by 'the original parameters'. (L. 194)

The original parameters are arbitrary parameters typical of the wind speed Weibull distribution we found in the area of study with the Lidar data. We used k=2.2 and lambda=8.5.

L. 262 More specifically, we chose Weibull parameters typical of the North Sea wind climate ($k = 2.2$ and $\lambda = 8.5$)

9. This could be re-phrased. In fact, I do not think a low number of samples can be called an advantage. (L. 207)

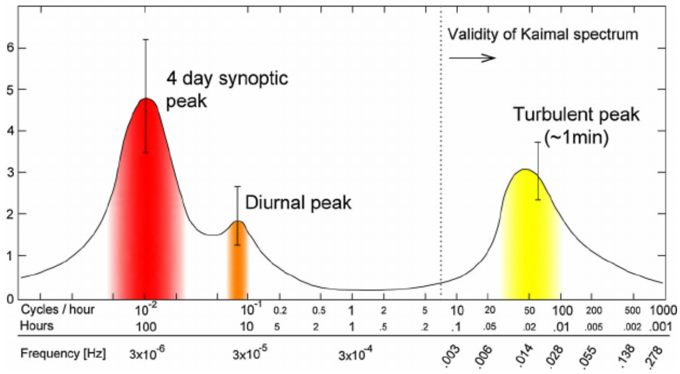
Corrected

L. 279 This limitation actually guarantees the statistical independence of measurements, nevertheless, since SAR satellites are on a sun-synchronous orbit, they pass always at the same times of the day, in the morning or in the evening. As a result, they cannot fully see the intraday variability of the wind.

10. I would like to see more information about these calculations - sounds a bit too good to be true. How many samples were used and did you calculate the ME or the MAE? (L. 215)

We do not see the reason why this result 'sounds a bit too good to be true'. We actually explained the reason why the error due to the SAR temporal sampling is expected to be low: 'It can be seen that the wind diurnal cycle is close to a 24 h period sinusoid. 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.'

To be more specific, we added a reference on the Van der Hoven spectrum of the atmosphere showing that there is a spectral gap between the diurnal peak and the small-scale turbulence. Therefore a sampling with a 12h time difference should indeed catch most of the diurnal and intra-day variability:



Regarding the errors, 'ME or MAE' are irrelevant here because we are not dealing with time-series, but scalar values. The errors are simply the error of the mean wind speed and the error of the wind power (in absolute value and %).

We added a table with the details of the results.

L. 283 The intraday variability of wind speed is low (Van der Hoven, 1957) and close to a 24 h period sinusoid (Figure 3). Therefore, since Sentinel-1 satellites pass at two possible times of the day separated by 12 h, according to the Nyquist-Shannon sampling theorem, it should be enough to capture the intraday variability. In order to verify this, we computed the mean wind speed and the extractible wind power using only Lidar measurements at 5 AM and 5 PM (UTC). Then, we compared results to the ones obtained using all Lidar measurements at any time of day. For all Lidar, the differences were found to be below 0.5% and 1%, respectively (Table 2).

Lidar	**Error of the mean wind speed in %**	**Error of the extractible wind power error in %**
HKZA	-0.34	-0.16
HKZB	-0.23	-0.01
LEG	0.36	0.94
EPL	-0.04	0.06
BWFZ01	-0.47	-0.08

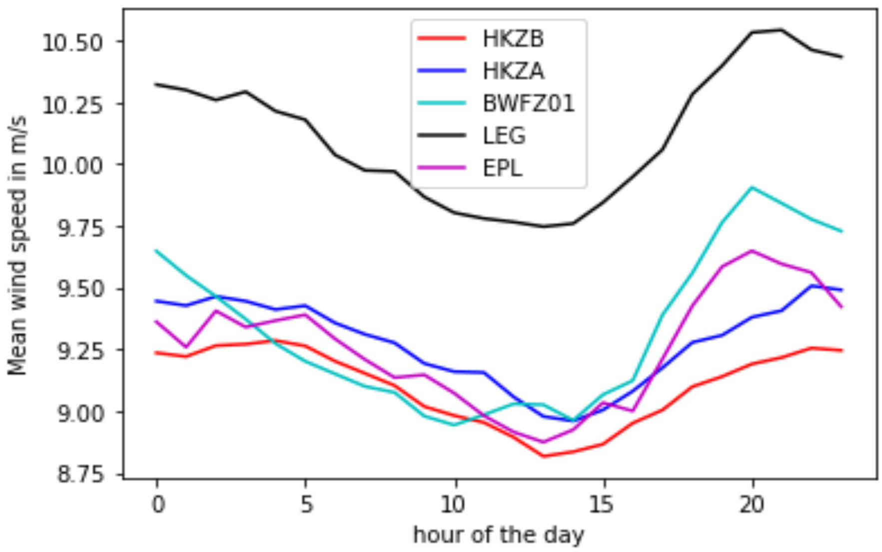
11. This sentence could be modified - it is not really about preventing the use of SAR data but rather about achieving the best possible accuracy on wind resource estimates. (L. 216)

Corrected.

L. 288 Therefore, the satellites are indeed able to capture most of the wind intraday variability.

12. Please comment on the differences between these curves somewhere in the text. Why is the diurnal variability less pronounced for HKZA and HKZB? (L. 221)

We agree that this difference is strange, but we are able to provide an explanation because we did not produce these data and do not have enough information about the measurement campaigns. It may be due to the lack of accuracy of the Lidars' first level. In the revised version, we give the curves at 120m instead and this problem seems to disappear.



13. The GMFs I know of are developed through triple-collocation using both model and in situ observations from buoys. Please check the literature and reconsider this sentence. (L. 221)

These GMF were designed mostly with ECMWF numerical model (see section 2 in Hersbach 2008 CMOD5.N: A C-band geophysical model function for equivalent neutral wind published by ECMWF). Actually, according to Stoffelen et al. 2017, the triple collocations with buoys were only used for validation purposes and a posteriori bias correction. Since this bias correction depends strongly on the considered scatterometers, in any case, we doubt it is relevant for SAR. So, we maintain our argument that the design of GMF is not well adapted to coastal areas because the ECMWF model is less accurate in these areas. However, we modified our statement to include your comment.

L. 60 Another reason is that GMFs were designed empirically using the ECMWF model as a reference, which may not be accurate in coastal areas (in-situ data were used only for validation and a posteriori bias correction, see Stoffelen et al., 2017, and references therein).

14. What is meant by 'interesting parameters'? (L. 245)

Corrected. We selected parameters known to be related to SAR errors due to physics o due to the retrieval algorithm design.

L.187 Regarding input parameters, we selected parameters related to SAR wind speed retrieval errors because of physics or because of the retrieval algorithm specificities.

15. When such a statement is made, we need to see the evidence - the numbers behind. Please provide them e.g. in a table. (L. 281)

Since we used a PBL more adapted to the higher boundary layer, we decided to verify the accuracy of the WRF surface levels. We did this by extrapolating Lidar data to lower altitudes, which is not very reliable, and found a strong bias. So we are unsure of these results and do not wish to present them, but, as a precaution, we decided to remove the WRF levels below 40m. If we had had accurate in-situ measurement below 40m, we would have given details, but we don't have any. Our aim here is just to tell the reader to be cautious with these surface levels when using a PBL adapted to higher altitudes, that we found a possible problem, but that we cannot conclude.

L.218 However, when assessing the WRF against Lidars, we found that the WRF wind speed had an unrealistic bias below 40m. It was unclear if this was due to the PBL adapted to higher altitudes, to a lack of accuracy of the Lidars at their first levels, or to the power law extrapolating these first levels to a lower altitude. In any case, as a precaution, we chose to use the WRF parameters at 40 m instead of the one from the surface level when producing the various input parameters.

16. If I understand correctly, the 10-m SAR wind is first modified through machine learning to match the lidar wind speed. Next, this 10-m wind speed is extrapolated up to 200 m. Since the starting point at 10 m is identical, it is not surprising that a good match between the wind profiles is found? (L. 292)

Separating the algorithm into two steps does not artificially improves the performance, because the training of the correction is done with the same training dataset as the as the one used to train the extrapolation. So the validation is independent and the starting point at 10m is not identical: the algorithm has to predict it before doing the extrapolation.

In any case, in the revised paper, we now transform the 10m SAR data into the equivalent 4m buoy wind speed with an algorithm trained with the NDBC buoy network located in the US. So the datasets are clearly independent.

17. It seems like there is some confusion about the term 'hub height'. For the 10 MW reference turbine used in this study, the hub hight is 119 m. You have used 200 m, which is approximately the hub hight + blade length i.e. the maximum height of the turbine. (L. 292)

See the answer to major comment 4 above.

18. Once again, it is difficult to follow this unless some evidence is provided in terms of numbers, tables, ... (L. 294)

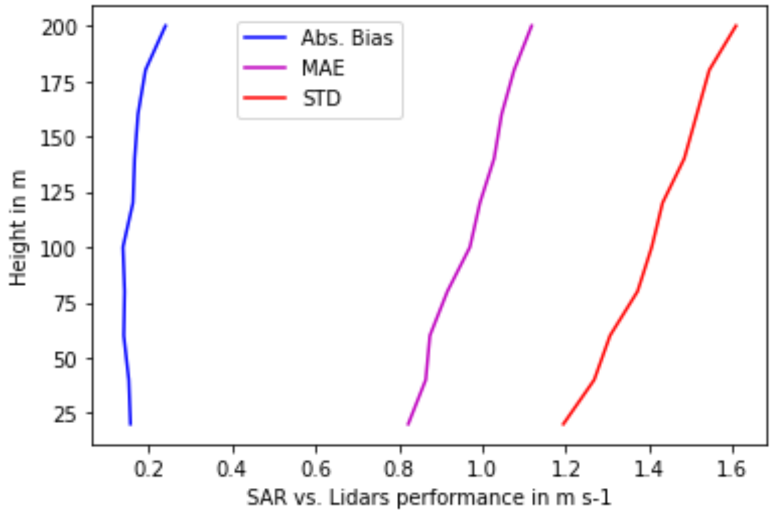
This statement was about a failed attempt, so it was neither necessary nor useful. We removed it.

19. What if the surface head flux from WRF is inaccurate as reported in the literature? How would this impact your results? (L. 296)

See the answer to major comment 5 above.

20. I suggest to put the 'altitude' on the y-axis as it is the convention in wind energy. And call it 'height' instead. (L. 301)

Corrected.



21. In the previous sections, many results were presented. I suggest to restructure so all results are presented in the 'Results' section. (L. 304)

Corrected. All machine learning algorithms performance were moved to the result section.

22. Why and when was this correction performed? Should be described in the Methods section. (L. 325)

Corrected. We gave more details in the Data section. See answer to minor comment 3:

Yes, actually we wanted to say that the numerical model was typical of the ones used by the industry, not the whole assessment. In any case, we were already using Lidar data to correct the WRF like the industry is doing (see the previous version of the paper in the results section). In the revised version, we explain now clearly this use of in-situ instruments to correct the WRF bias in the Data section.

L.141 Moreover, since the WRF is typical of numerical models currently used by industry, we also used it as a reference to assess the benefits of using SAR data (Section 4). Since numerical models are often combined with in-situ measurements to increase their accuracy, we also corrected the WRF bias. The extractible power estimated by the WRF was found to be underestimated by 3% compared to Lidars.

23. I think, the most striking result is that the coastal wind speed gradients are resolved by SAR and not by WRF. Please elaborate on that. (L. 347)

Corrected.

L. 375 In particular, the coastal wind speed gradient, which is often crucial in offshore site assessments, is resolved by the SAR and not by the WRF (see Figure 12).

We also added a figure showing a coastal gradient on a perpendicular to the shoreline.

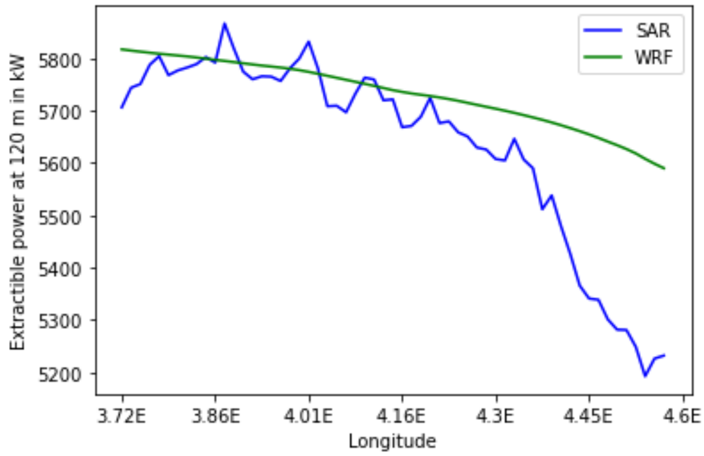


Figure 12: Extractible wind power coastal gradient at 120 m on a horizontal line at the top of Zone 2 estimated by the WRF and by SAR satellites.

24. This belongs to the 'Methods' section. Please elaborate on the Koch filter or at least give a reference. (L. 352)

Actually, the use of Koch filter was already explained in the Data and Methodology section about SAR data (now 2.2) and a reference was given.

25. But mast/lidar observations are needed as well? This should be discussed in terms of the practical application of your method. (L. 380)

The measurements used in the validation are statistically independent since they are measured with more than 48h time difference, so we do not think mast/lidar are needed to apply the method. Moreover, the revised version now uses a round-Robin validation, which shows clearly that that the method can be trained in one place and applied in another.

However, we are not satisfied with the round-robin validation since the Lidars are too close from each other, and because it was not possible to fully test the approach in other seas due to the lack of freely accessible Lidar data. Therefore, we would like to perform more validation with more Lidars in the future.

26. Is it realistic to develop a general approach for all seas? Or will there always be a need for in situ measurements? (L. 385)

As explained above, in-situ measurements are not needed. However, since the training of the extrapolation was done in the North Sea, we do not think that it can be applied directly in all seas. To apply the method in seas having a very different wind climate, like the Mediterranean Sea, we expect that Lidars located in the region would need to be used to train the algorithm. We clarified it in the conclusion of the paper.

L.430 Further research should focus on removing remaining artefacts on the SAR wind power maps, such as swath edges, bright targets, and the effect of bathymetry. Moreover, since the *method was validated only using Lidars located in the North Sea, the extrapolation algorithm may not be adapted to meteorological conditions in seas having a different wind climate. In that case, wind profiles measured by Lidars located in the region where the site is located would need to be included in the training dataset and used to validate the method.*