Response to Referee #1

Daniel Hatfield, Charlotte Bay Hasager, Ioanna Karagali

February 14, 2023

First of all, thank you very much for taking the time to review this article and for your positive comments and suggestions. Below are responses (in black) to all of the referee's comments (in blue).

General Comments

"The paper describes the application of machine learning algorithms to vertically extrapolate near surface winds derived from satellites data. The focus is offshore, in Northern Europe. The paper is generally well written (although some grammar improvements here and there would be recommended), and it represents a nice academic application of machine learning to the wind energy sector. However, I question the practical utility of this approach in real world applications. The main reason for this is that satellite-derived observations of wind speed at a given location are only available twice a day, and always at around the same times. I am really struggling in finding a situation where someone would be interested in knowing hub-height wind speed only at these two hours of the day. While practical applicability is not a strict requirement for having a paper published, I still think this limit should be at the very minimum highly stressed in the paper, and language softened to reflect the limited applicability of the results described in the analysis."

Thank you very much for you comments and attention to detail while reviewing the manuscript. We agree that the practical application of the methods presented in the study are limited, but would like to remind the referee that this was a *proof of concept* style of paper, exploring the potential of using machine-learning method as a way of extrapolating satellite data as opposed to the alternatives that have not had much success. This was done in hopes to further utilize the invaluable daily satellite surface wind coverage for longer-period wind resource assessment.

We do agree that the language used in the paper needs to be "softened", in which sections and word-choices have been changed to stress the limitations/potential of applying machine learning methods.

Specific Comments

1. L.19: add "above the sea surface" or something similar. Added

2. L. 26: 275 m seems like a very specific threshold – can you provide a reference? We have included references for both ferry and buoy based lidars with heights up to 275m in line 26 as (Rubio et al. 2022 and Hatfield et al. 2022)

3. L.45: "predicting" instead of "predict" (or "to predict") Added

4. L. 66: "higher" instead of "greater" Added

5. L. 106: in the text you mention water temperature, while in Table 1 you mention sea surface temperature; please be consistent. This has been changed in Table 1 to Water Temperature

6. L.111: do you mean that all variables have exactly an availability of 85%, or greater than 85%? We have clarified this in line 113 "All measured quantities show a data availability above 85% with the exception of WT (76%)"

7. L.123: what's the temporal resolution? The temporal resolution is *daily* coverage (which is now clarified in the text in line 121). With the high heat capacity of water small diurnal changes are observed and uncertainties are provide by Høyer & Karagali 2016 (also mentioned in the text). It should be noted that the SST product has a mean difference of -0.06°C compared to moored buoys and a 0.46°C standard deviation of the differences.

8. Consider moving all the details regarding data access to the 'Data availability' section towards the end. Thank you for the suggestion, we will however include the data in the text to remain consistent with other papers using the satellite/FINO data.

9. Figure 1 caption: "shapes" instead of "rectangles". Added

10. It is hard to fully understand Table 3 (and some of the discussion in this section) without a clear explanation of the temporal frequencies considered here for the various data sources. Does "total data" refer to 30-min average time periods? And "Concurrent data with ASCAT" to 30-min average time periods in which an ASCAT data point was recorded? Please clarify in the paper. This is correct, "total data" refer to 30-min average time periods at the masts, "concurrent data with ASCAT" refers to 30-min average time periods in which an ASCAT data point was recorded. This is now updated in the Table 3 label.

11. Table 2: "FINO" is repeated twice in the left column. Removed

12. Why using the cosine of wind direction only (and not the sine, too)? We have used both at one point in the early stages of this work where using the cosine of the wind speed performed better in the ML training (resulting in lower RMSE).

13. Once again, being clear about the temporal resolution of the data used is key to understand whether the random split between train and test sets is a right choice, or auto-correlation effects might play a role in artificially enhancing the ML results. Added in Table 2 caption as: "All of the data measured from the FINO masts are 30 minute averaged."

14. Table 5: I would argue that the most relevant comparison is ML vs NEWA at each site, so I would suggest adding a horizontal line after each site, and highlight in bold the "winner" metrics at each site. This is a very good suggestion and is now incorporated in Table 5 along with the addition of the same comparison with NORA3 at FINO for the same periods as NEWA.

15. L.170: clarify that the error values refer to the test set. This is clarified in line 171 as "...with the test dataset..."

16. L.173: "Note" instead of "not". Added

17. Section 3.2: please clarify how the mean wind profile from the RF was computed. Do you simply apply the RF to the whole period, and average results? Or to the test set only? This is now clarified in the text in line 172: "The model trained in Section 3.1 is applied to the entire 12-year collocated dataset at all heights at FINO3 from 31 m to 107 m."

18. Section 3.3: why not including FINO2 as well? FINO2 is much too far away and the Baltic sea has a very different marine atmosphere. As this is still an article exploring the *proof-of-concept*, we have limited the spatial extension / round-robin approach to the North Sea. It should also be noted in the ML papers extrapolating wind speeds (i.e. Optis et al. (2021) and Bodini et al. (2020)) that use the round-robin approach, their distance are less than 100km apart, whereas the distance between FINO1 and FINO3 already exceeds that of previous work (136km). Thus, the comparison with FINO2 is beyond the scope of this article.

19. Section 3.4: can you somewhat verify your hypothesis of horizontal homogeneity by looking at spatial variability of the meteorological variables from NEWA and the reanalysis product? In Figures

4a and 5a we can see the spatial variation of both the wind speed (from NORA3) and SST (from DMI L4 SST) both with very low variation across the 125km^2 area. The FINO3 area is also in open ocean, far from the coast and without islands within the study area. We think this is a very fair assumption to make. We have also verified using ERA5 data for the North Sea that there is small variations in air temperature (± 1°C similar to that of the SST variation) across the chosen area which is not shown.

20. Figure 4: once again, more clarity is needed when explaining what is being plotted and described. What's the temporal extent of what is shown? Are NORA data taken only at time stamps at which ASCAT data are available? How about the ML-extrapolated winds? Are we really comparing apples to apples? Once that is clarified, please adjust text accordingly (it is misleading to state you are showing 2018-averaged winds, if you are only cherry picking time stamps). ASCAT is the limiting feature, all time-stamps need to be concurrent with ASCAT. This is however is not the case in Figures 4 & 5, these included yearly averaged 30 minute NORA3 grids and have now been updated to be concurrent with ASCAT so that we are comparing "apples to apples". The Figures 4 & 5 are now updated to include concurrent NORA3 data instead of yearly averaged for 2018.

Thank you very much for the comment!

21. Figure 4: why wasn't NEWA included? Cheynet et al. (2022) has shown that NORA3 out-performs the NEWA at FINO1 but this is not mentioned until the Discussion. A clarification is now added in lines 228 "It should be noted that only NORA3 will be included in the spatial comparison with the RFM as it has out-performed NEWA at the FINO3 mast in Table 5 and in Cheynet et al. (2022) at FINO1.". NORA3 comparison was also added in Table. 5 to further this argument.

22. Conversely, why wasn't NORA3 included in the earlier analysis (Table 5)? It is essential to know how well it compares to observed winds in order to use it as proxy for the truth here. Added

23. Figure 6: please change labels and instead list ALL the satellites available in each time period. Also, specify that this is a cumulative plot. What do you mean by "concurrent" in the caption? Concurrent is confusing, it is now changed to "Cumulative number of samples of ASCAT observations at the FINO3 location from 2010-2022. The vertical lines represent the launch of each MetOp satellite as well as the decomission date of MetOp-A." in the Figure 6 label. The Figure is now updated to show which satellites were also in operation during this time period.

24. Discussion of Figure 10 is key (see my main major comment above), and in my opinion should be moved way earlier in the paper. I agree that this is a major point of discussion, but we do think it is in the right place within the results. This paper explores the implications of the use of ML on satellite extrapolation where we explore vertically (wind profile), horizontally (round-robin) such as previous work but also the spatial (in comparison with NORA3) and temporal (in sampling) domains due to the nature of satellites. We completely agree that the discrete nature of the satellites is one, if not, the largest limiting factor within this paper, however we think that is explored in the discussion. Whereas, the layout of the paper slowly builds on expanding the simple vertical extrapolation of the satellite at each FINO mast.

25. Figure 10 caption: specify when referring to bars vs lines. Also, specify you are referring to local time and not UTC time (I believe). All times throughout the paper have been recorded in UTC (ASCAT, FINO, SST, NORA3) which we have added in the Data section in line 146 as "all data used from all sources is recorded in Coordinated Universal Time (UTC)". But we have added in the Figure for clarity and clarified the bars/lines.

26. L.332: decreases instead of increases? Yes, changed.

27. Data availability: why not sharing the model algorithm scripts as well? Unfortunately, there is no intention from the authors to publish the scripts as well.

28. Please double check references and make sure each has a DOI. Done