In this document, the reviewer's comments are in black, the authors' responses are in red.

We thank the reviewer for their thoughtful comments, which gave us an opportunity to revisit our analysis.

The manuscript describes a statistical (machine learning) approach to derive the uncertainty of wind data from mesoscale model simulations for longer time periods. The topic is relevant for wind energy. Thus, the manuscript should be published after revisions.

The manuscript is clearly written. Nevertheless, reading the manuscript rises a number of issues and questions which should be addressed in more detail in the manuscript.

(1) Lines 37-39: Using marine tower data (e.g., available offshore off the coasts of the Netherlands, Germany or Denmark) should at least be mentioned in the introduction in parallel to the possibility of using floating lidar data. There is more offshore wind data than those from buoys. Please add some references.

We have rephrased this part as: "Direct observations of the wind resource offshore are oftentimes limited to buoys, which offer measurements at very limited heights. Hub-height measurement of the wind resource offshore can be achieved with either offshore meteorological towers (e.g., Neumann et al. (2004); Fabre et al. (2014); Peña et al. (2014); Kirincich (2020)) or floating lidars (Carbon Trust Offshore Wind Accelerator, 2018; OceanTech Services/DNV GL, 2020). However, the often prohibitive costs connected to both these measurement solutions limit their availability to a handful of locations, despite recent efforts in leveraging their punctual hub-height measurements for wind speed vertical extrapolation over a larger region (Bodini and Optis, 2020; Optis et al., 2021a)."

(2) Given that there is some tower data (some of them for more than ten years), could testing of the results from this study against long-term time series from the FINO1 or FINO3 platform in the German Bight be an option?

Absolutely, using long-term tower data could be used in future work to model the actual error between WRF-modeled and observed wind speeds at hub-height. We had already mentioned this opportunity in the conclusions, but only with reference to floating lidar observations. We have now added an additional sentence to include tower data, too. That part now reads: "While floating lidars with publicly available data have only been deployed in the California OCS very recently (Gorton, 2020), a few lidars have been deployed off the U.S. Eastern Seaboard for more than 1 year. Observations from long-term offshore meteorological towers are also available in the North Sea in Europe. Our analysis could be expanded by first comparing the model-related boundary condition and parametric uncertainty with the full modeled wind speed uncertainty calculated by comparing modeled data and observations. Then, our proposed approaches could be expanded to temporally extrapolate the full modeled wind speed uncertainty; for example, quantified in terms of the variability of the residuals between modeled and observed wind."

(3) The addressed wind conditions in this manuscript are those in coastal areas. Therefore, the reviewer is wondering why coastal effects in the wind fields are not addressed. This should at least be mentioned in the Introduction.

We have added the following part to the Introduction: "Given these constraints, numerical weather prediction (NWP) models at the mesoscale are often used to obtain an in space and time continuous mapping of the available offshore wind resource at the heights relevant for commercial wind power plant deployment (e.g., Mattar and Borvarán (2016); Salvação and Soares (2018)), with some studies (Papanastasiou et al., 2010; Steele et al., 2013; Arrillaga et al., 2016) also focusing on the validation of modeled coastal wind effects, such as sea breezes, which have a significant impact on offshore wind energy production (Archer et al., 2014). "

(4) Some configurations of the mesoscale model WRF have problems representing the atmospheric boundary layer in the transition region between land to open sea (see, e.g., Siedersleben et al., 2018, DOI 10.1127/metz/2018/0900). Has the WRF configuration chosen for this study been checked for the ability to properly simulate the transition from land to sea?
We did not specifically check this, but wind flow in the region is almost entirely from the northwest, with large open fetch and minimal land-sea interaction. Therefore, we expect any impacts from improperly modeling the transition layer would be minimal in this context.

(5) This study addresses uncertainties in wind speed. How does the magnitude of these possible uncertainties in wind speed modelling relate to, e.g., interannual wind speed changes due to long-term oscillations in the global circulation (NAO etc.) or due to climate change?
Assessing the impact of climate change and/or long-term oscillations is strictly outside the scope of our manuscript. However, to guide the interested reader in this analysis, we have calculated the interannual variability in modeled 100-m wind speed from the 20-year record used in the analysis, and included the following map in the Supplementary Materials:
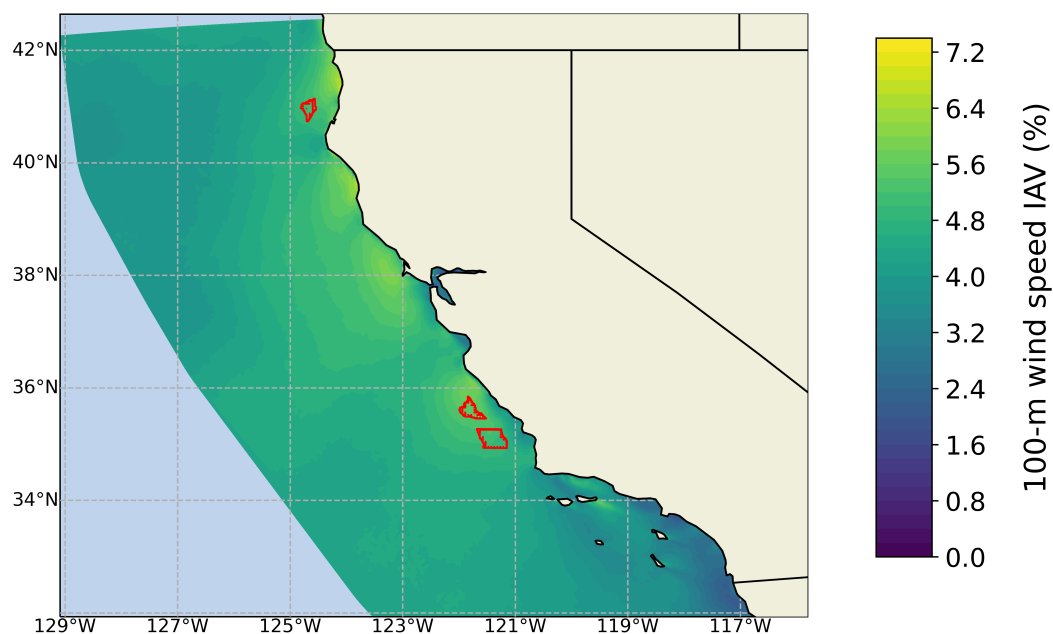

Also, we have added the following sentence at the end of the Conclusions:
"Finally, future work could focus on how interannual wind speed variability caused by climate change or long-term climatic and atmospheric oscillations (e.g., the North Atlantic oscillation) compares with the quantified uncertainty in modelled wind speed and how that should be taken

(6) Figure 7: Why do unstable thermal stratification conditions have such a small share of all cases? Shouldn't the lagged annual SST variation (due to the much larger thermal inertia of the water) compared to the air temperature variation lead to roughly equally frequent stable and unstable situations (stable in late winter and spring, unstable in late summer and autumn)? Please give an explanation for the found bulk Richardson number distribution.

To justify the distribution of bulk Richardson number, we have calculated the difference between air temperature and sea surface temperature at the considered site, using the 20-year WRF data:
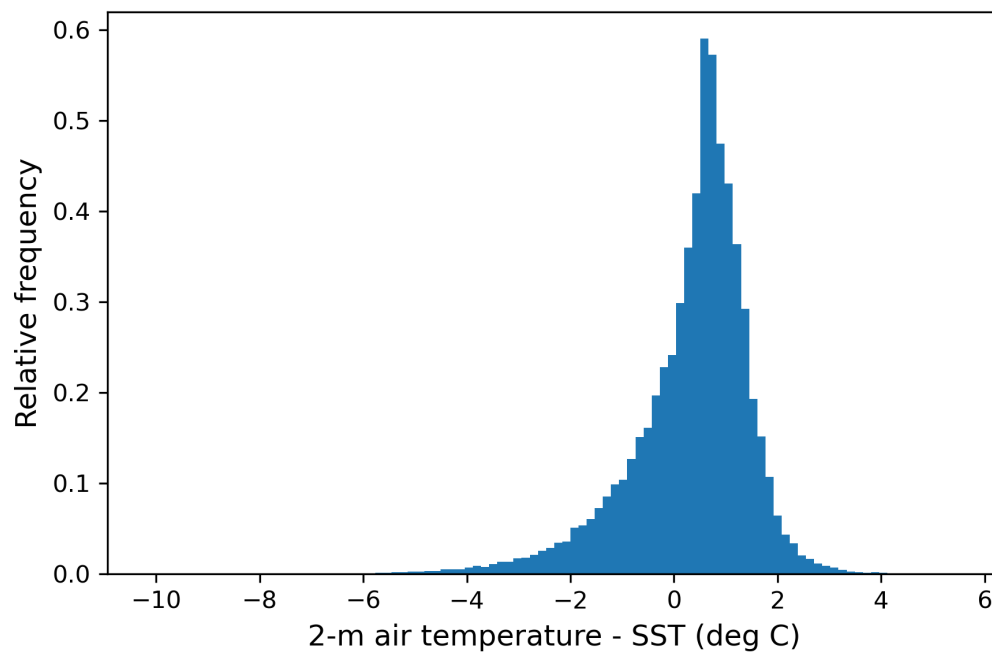
As the histogram shows, the air temperature is generally higher than the sea surface temperature, likely because of ocean upwelling in the region, which leads to a generally cold sea water. This induces a stable stratification, which justifies the distribution of bulk Richardson number found and reported in the paper. We have added a comment on this in the paper, too ("This is consistent with the sea surface temperature being generally colder than the near-surface air (because of ocean upwelling), which causes a predominantly stable stratification.").

(7) Why is the 2 m inverse Obukhov length chosen as a parameter at all in Section 2.2? This parameter can only be relevant for the whole rotor area of the future wind turbines in cases with perfect vertical mixing. In stable conditions the depth of the Prandtl layer is much shallower than turbine hub height so that the near-surface Obukhov length is most probably irrelevant for the 100 m wind speed. This is likely to be also true for cases in which internal boundary layers are present (which frequently happens in coastal areas).

See our answer to comment #9.

(8) Figure 3 demonstrates that the 2 m inverse Obukhov length is not a meaningful parameter, because it is always close to zero. Logically, the weight 0 has been attributed to this parameter in Table 4. So, once again, why has this parameter been included into this study?
See our answer to comment #9.

(9) Assumed that the inverse Obukhov length has nevertheless some relevance, a not-discussed contradiction is that there seems to be a slight bias of the inverse Obukhov length towards instability in Figure 3 while there is a strong bias of the bulk Richardson number towards stable conditions in Figure 7? This should be explained. It seems to confirm the irrelevance of the inverse Obukhov length stated above.
The observed range of inverse(L) is actually a quite typical one, if we consider that a value of $L =$ +- 200m (a somewhat typical threshold between neutral and stable/unstable conditions) would translate in a 1/L threshold of +-0.005 m^{-1}. We have changed the x-axis limits in figure 3 to make the plot (where no real bias emerges) easier to understand.

That being said, we have performed the feature importance calculation for the input parameters to the ML algorithm at one test site (one of the three wind energy lease areas), and actually found how inverse(L) was the second most important input feature for the algorithm:
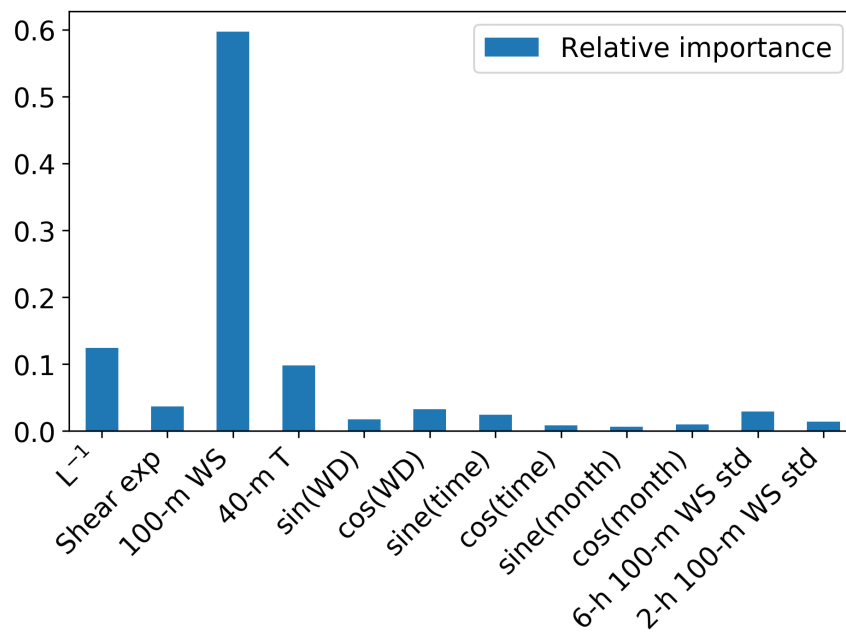
In any case, we agree with the reviewer that different choices for the stability parameter are possible, and using a different one could potentially improve the performance of the proposed algorithms (as hinted by the fact that inverse(L) was not deemed useful by the AnEn algorithm). We have added the following comment to Section 2.2: "We also acknowledge how different choices for the atmospheric stability parameter could be explored, potentially leading to a more accurate representation of stability at the heights of interest for wind energy development compared to the near-surface Obukhov length."

(10) A major finding of this study is the increase in wind speed uncertainty close to the coast and for non-neutral stratifications. Does this point to weaknesses in the used method or to weaknesses

in the used mesoscale model? Or is this a natural feature which comes out of this study? Please discuss this point in the Conclusions.

The larger uncertainty in modeled hub-height wind speed near the coast was also found in Optis et al. 2020 when looking at the WRF ensemble members only. In their report, they comment that "Most of this sensitivity can be attributed to the choice of PBL scheme (MYNN or YSU) and the diverging wind profiles associated with each under strong stable atmospheric conditions near the coastline." We have added the following sentences in the Conclusions of our paper: "In general, we find that the offshore wind speed boundary condition and parametric uncertainty increases near the coast. While the accuracy of the AnEn approach significantly degrades near the coast, the larger values in hub-height wind speed boundary condition and parametric uncertainty near the coast were also seen from the variability among the WRF ensembles (Optis et al. 2020), and attributed to diverging wind profiles associated with the choice of PBL scheme under strong stable atmospheric conditions near the coastline.".