Reply to Reviewer #3:

This study is clearly presented and well written. The objective is to improve upon the recent work of Voerman et al. (2020) that attempted to invert near-surface wind speed and wind direction from ocean wave buoy datasets provided by the NDBC network of coastal and offshore buoys. That previous study provided a thorough review of wind-wave interaction as it pertains to buoy measurements and this inversion. The present study bypasses the geophysical basis and instead focuses on a sort of brute force neural network (DNN) approach to the wind estimation task using the NDBC data archive of five freq. dependent Fourier coefficients that are used to approximate the directional gravity wave spectrum from long to intermediate scale surface waves (both swell and wind sea). The study appears to use data from the entire buoy station network to develop separate wind speed and direction algorithms, provides detail on the network training and several relevant DNN adjustments during the training process, and then results that show some promising capability to provide wave-buoy derived wind estimate that agree better with the buoys' anemometer measurements. They also find that the winds derived in this manner appear to lag behind the actual surface winds in time by 30-60 minutes - and thus their final algorithm estimates not the wind at the present time, but actually the wind that occurred one hour before. They also find, as did the recent Voermans et al. study, that their best algorithms still have limitations at lower and higher wind speeds where the wave information does not unambiguously relate to the wind.

The author would like to thank the reviewer for the patience for reading the entire paper carefully and the encouragement. The comments from the reviewer are very helpful for the improvement of the study. Some revisions are made to the manuscript according to them. For few comments the author has different opinions on, explanations are given in this reply. The author hopes the revised manuscript is acceptable for the reviewer.

While this paper does show some potential for a neural network algorithm that takes the basic directional wave information provided by NDBC and outputs wind information, it does not appear to move things too far forward from the Voermans study they follow on from and the low and higher wind speed regime limitations that were highlighted in that study. What it does illustrate is that a DNN can improve on the semi-analytical approach used in the previous investigation.

The author admits that this study does not move things forward from Veormans et al. (2020) with respect to the geophysical basis of the wave spectrum-based wind-estimation model. However, the final aim of establishing such a model, in the author's opinion, is to have the ability to estimate the wind information as accurately as possible. Since the underlying physics and the possibility of establishing such a model have been discussed by Veormans et al. (2020), this study focuses on the improvement of accuracy.

Because the relationship between inputs (spectrum) and outputs (wind) can be highly nonlinear and there might be some 2^{nd}-order effect that is difficult to be considered in the semi-analytical model, the author simply used the DNN model to "learn" the input-output relationship to obtain better accuracy. The author believes that DNN is the best suitable for such problems: we have some understanding of the relationships between inputs and outputs, but the detailed physical model is too complicated to establish analytically. The results show that this selection is not bad, the accuracy of the estimated wind is improved significantly from Veormans et al. (2020) in conditions of moderate wind speed (the overall RMSE for 3-20 m/s wind speed is improved from ~2 m/s to ~1.2 m/s without time delay and ~1 m/s with a 40-minute time delay).

Regarding low and higher wind speed regimes, the author believes that this is the problem of almost all indirect wind-estimation models and one of the challenges of almost all wind measurement technologies. For low wind speeds, the response of surface waves is too weak while the impacts of other geophysical noises might be strong. For high wind speeds, the air-sea interaction is complicated while we do not have sufficient samples (there are less than 100 samples for U10>21 m/s) to build a robust model. Still, compared to Veormans et al. (2020), the DNN model also performs slightly better in high and low winds. For example, the RSME for 1 m/s, 2 m/s, 3 m/s, 15 m/s, 17 m/s, and 20 m/s buoy wind speeds were all improved significantly: 3 m/s→ 2 m/s, 2.5 m/s→1.2 m/s, 2 m/s→1 m/s, 2.5 m/s→1.5 m/s, 3 m/s→2 m/s, and 4 m/s→3 m/s, respectively.

The finding that there they is an apparent delay between the wind speed and the wave-inferred wind speed is not physically inconsistent with Voermans et al. (2020) Figure 9g where the wind acceleration is related to model error residuals. However, there is an additional issue for the authors to consider first. The wave buoy measurements provided by NDBC have a center time that is 30 min past the top of the hour with data collected +-10 min of that time. The authors do not clearly provide detail on the NDBC wind products they are using, but if that product is the stdmet product then the center time for that 8 min. avg wind estimate is at minute 46 (measurements made from 42-50). Thus there is an inherent 15 min offset with the hourly wave data leading the wind. This factor may also color why the previous wind measurement is more highly correlated with the wave-inferred winds. Finally, the NDBC network does contain a large number of continuous wind measurement buoys where winds are measured every 10 minutes. Thus the authors have the opportunity to investigate the actual lagged correlation between DNN wave-derived winds and the anemometer data with 10 min resolution and perhaps at varying wind speeds.

The data used by the author is the archived data from National Centers for Environmental Information (https://www.ncei.noaa.gov/data/oceans/ndbc/cmanwx/) in NetCDF form, and the actual acquisition time of wind, waves, and continuous wind are provided separately using different dimensions. However, the suggestion from the reviewer, using continuous wind to investigate the lagged correlation, is very helpful. This work was conducted in the revised version of the manuscript, and the result is shown in the new Figure 3 (Also shown here as Figure R1). It is found that the overall best error metrics for wind speed and wind direction were found at 40-50 minutes and 40-60 minutes before the end of the waves' end sampling time.

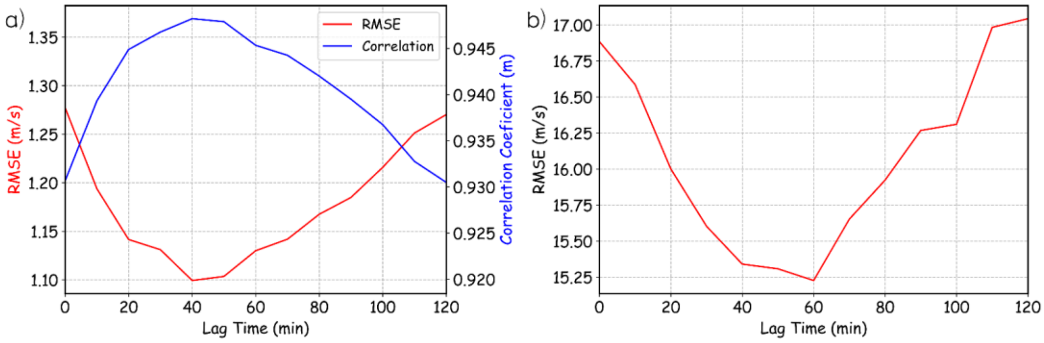

Figure R1. Figure 3 in the revised manuscript: (a) The RMSE and CC of the DNN-estimated wind speed as a function of lag time between wave and wind measurements (waves' end sampling time minus winds' end sampling time). (b) The RMSE of DNN-estimated wind direction as a function of lag time between wave and wind measurements for wind speed higher than 7 m/s.

Following the suggestion of the reviewer, the author also investigate the lag correlation at different buoy wind speeds, and the results are shown in Figure R2. For different wind speeds, the best correlations (minimum RMSE) for wind speed were all found at the time offset of ~40 minutes (the lag for U10>12 m/s is not significant). Therefore, using a simple offset of 40 minutes should be sufficient for the model. Based on this result, the DNN models were retrained using 40 minutes delay.

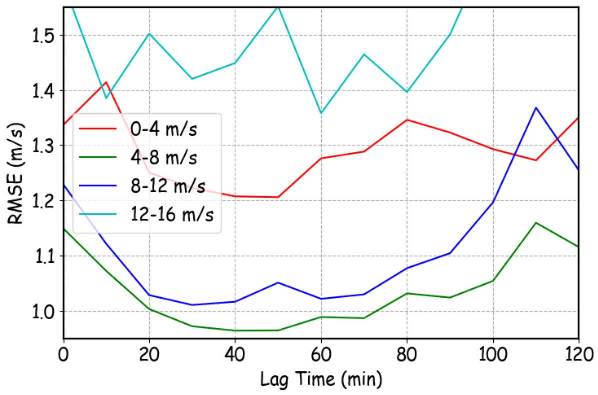

Figure R2. The RMSE of the DNN-estimated wind speed as a function of lag time between wave and wind measurements for different wind speeds.

Regardless, this issue points outs that using a series of DNN models to sort this out is an indirect and poorly-posed reverse engineering approach to infer the growth or dissipation rate of wind waves, as well as an illustration of the fundamental limitation in the use of surface waves to provide accurate wind measurements under a full range of wind forcing and sea states discussed in Voermans et al. (2020.

Yes, the author thinks this might be the potential problem of all methods based on artificial intelligence: it is difficult for them to directly tell us new physics. And with such a large number of samples (million level), the DNN model has probably reached the limit of estimating wind using surface wave (at gravity range). If this model cannot solve the high/low wind problem, probably neither can other models, unless we have more samples of extreme wind or have a wider range of high-frequency tails (probably also not very helpful as frequency spectrum at the tails is very strongly impacted by surface current).

A significant concern related to this time delay is the need to explain the potential implications of their DNN-derived estimates for users such as forecasters. The final DNN models are tuned to give wind speed and direction from the hour before. Thus I believe the first sentence of the Concluding Remarks should clarify this point. I believe the authors should consider a revisit of this product. Perhaps they should provide statistics and models for two wind options, the nearest time wind and the previous hour winds?

The author has revised the second sentence of the Concluding Remarks to "…DNNs that can estimate U10 and wind directions ~40 minutes ago from high-frequency wave spectra…", which should be clear now. Regarding the two options, the statistics of the nearest time wind model are shown in Figure s2a and 2d. However, the author did not emphasize the "nearest time wind" option for three reasons: 1) The data of one hour's delay (now only 40 minutes' delay) can already be regarded as near real-time, which can be very useful for the operational application such as forecast. 2) In fact, the DNN model to estimate "nearest time wind" also has a better agreement with the wind 40 minutes ago. Therefore, even if the application is very sensitive, the 40-minute-delay wind can be directly approximated to the "nearest time wind" with a similar accuracy to an ad hoc model. There is no need to use two models.

The model sensitivity tests in the discussion section are an ad hoc revisit of the more in-depth work of Voermans et al. (2020) and previous work (e.g. Jusko et al., J. Phys Ocean. 1995). But simply withholding part of the frequency spectrum from the inputs does not provide new results. It confirms, as the authors note (lines 205-210), what has already been shown in terms of the importance of the higher frequency portion of the spectrum closer to the wind sea peak frequency and the tail of the spectrum. The authors appear to perform this test in the same way for all wind speeds and conditions and perform the RMSE assessments similarly for all winds. This is a course sensitivity test. Perhaps something more creative could be done to investigate the potential to modify inputs with a goal to improve performance at low and high wind speeds?

There are also two reasons to do the sensitivity test. One is simply to refine the input of the model. The DNN was established in a very brutal way of including all Fourier coefficients at all frequencies as the input. Using such a sensitivity test can let us know which of them do contribute to the wind estimation. This will help us to make the size of the DNN smaller so that can be more easily trained. This sensitivity test also tells us that including the r_1 information (which describes the directional spreading for each frequency) is helpful for the estimation of wind speed probably because the directional spreading of high-frequency waves also contains the information of wind speed. In fact, the author also tried to establish a DNN model for U10 estimation with only wave spreading information (r_1 and r_2), and the resulted overall RMSE can also reach 2.2 m/s, as shown in Figure R3. Therefore, such a simple sensitivity test can still provide some new information.

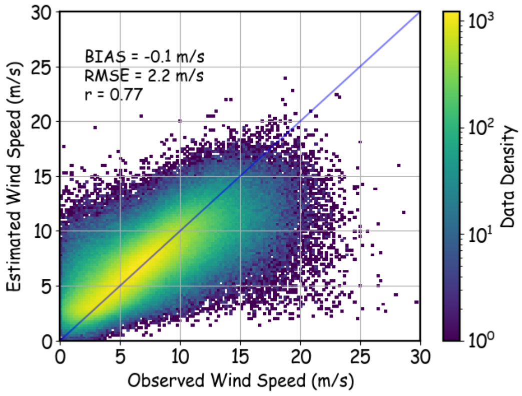

Figure R3. Scatter plot of collocated DNN-estimated wind speed using only wave spreading information (r_1 and r_2) as input and direct-measured wind speed.

The other aim is to check whether the modulation of low-frequency waves on high-frequency waves has a significant impact on the model. Previous studies have shown that the modulation of low-frequency waves on capillary waves can be a 2nd-order factor for wind remote sensing (e.g., Stopa et al. 2016, Li et al. 2018, Jiang et al. 2020). However, the results in Figure 5 (original Figure 4) show that this modulation effect is not important for wind estimation from surface gravity waves.

Reference:

Jiang, H., Zheng, H., and Mu, L.: Improving Altimeter Wind Speed Retrievals Using Ocean Wave Parameters. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 13, 1917–1924, doi:10.1109/JSTARS.2020.2993559, 2020.

Li, H., Mouch, A., and Stopa, J. E.: Impact of Sea State on Wind Retrieval from Sentinel-1 Wave Mode Data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.,12, 559-566, doi: 10.1109/JSTARS.2019.2893890, 2018.

Stopa, J. E., Mouche, A., Chapron, B., and Collard F.: Sea state impacts on wind speed retrievals from C-band radars. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 10, 2147–2155, doi:10.1109/JSTARS.2016.2609101, 2017.

More detailed information on the specific wind and wave buoy products that they used in training, their data filtering and quality control, and references describing the approach that NDBC uses to extract the directional wave Fourier coefficients should be provided.

The data used in this study is the data archived in National Centers for Environmental Information, so that the data has been quality controlled by NDBC. The author did not do too much quality control for the data except for removing the records with bad-quality flags. More detailed information on the data products was provided and the corresponding reference of measuring Fourier coefficients (Steele et al. 1998) was also provided in the revised manuscript.

Reference: Steele, K. E., Wang, D. W., Earle, M. D., Michelena, E. D., and Dagnall, R. J.: Buoy pitch and roll computed using three angular rate sensor. Coast. Eng., 35, 123-139, 1998.

Given what is observed in terms of data quality in the section surrounding Figure 3, is there any concern that such corrupt data are present in the training and/or validation datasets? Moreover, as noted in the next paragraph, it would seem to be obvious that the algorithm training set should not include buoys where there is strong known wave/current interaction such as 46087 and 46088. This would be a highly unusual case of wind-wave-current interactions that would not be desired in a general-purpose wind algorithm that only uses the 5 Fourier coefficients and no surface current data as inputs.

The corrupt data are present in the training and validation set. However, the number of samples for these corrupt data is very small which can neither impact the training of the model nor the validation of the model (with respect to error metrics). Therefore, the author did not re-train or re-validate the model for simplicity.

Similarly, regarding the cases with strong currents, because the current data is not available from the buoy, it is difficult to remove the cases with strong currents. Even for the buoys 46087 and 46088, the currents are not always strong. Including them in the training/validation dataset will also have almost no impact on the results (Figure R4). Meanwhile, when the users are using this model, it is also difficult for them to know whether there are strong currents at the location of wave measurement. Therefore, the current is considered as a source of environmental noise for this model. The model is not inapplicable in conditions of strong currents, but the accuracy will slightly decrease.

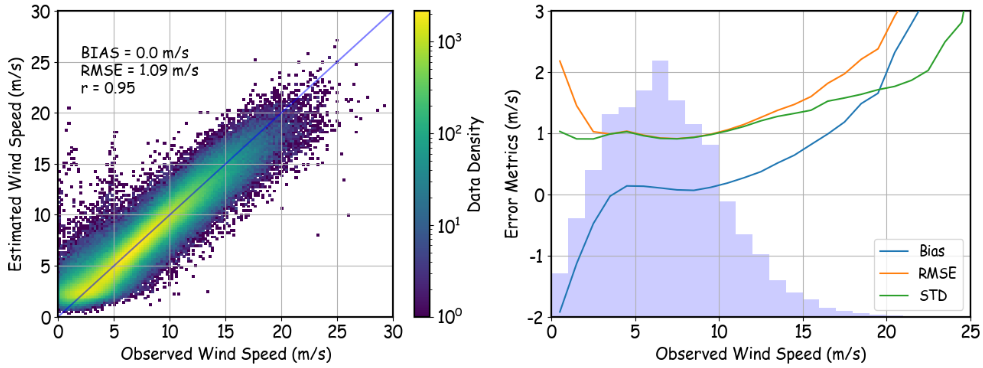

Figure R4. The same as Figure 2b and 2c in the manuscript, but the data from buoy 46087 and 46088 is excluded from the training and validation dataset.

Similarly, was there any consideration given to differentiating between coastal, offshore, and/or differing wind-wave climate buoys in the model input training sets to improve performance, for example at low or high wind speeds.

In fact, the author has not only tried to differentiate the coastal and offshore conditions, but also tried to use the buoys' distances to the nearest coast as an input term of the DNN model. However, this consideration did not improve the model so that the author did not mention it in the manuscript. However, on the bright side, this also indicates that the generalization ability of the DNN model is good and the users do not need to deal with several models for different conditions.

Different wind-wave climates of buoys were not considered in the model. But differing the location of the buoys has some implications of the wind-wave climate. The author even tried to establish a DNN model for each buoy, which did not improve the model, either. According to the suggestion of the reviewer, the author also tried to using the climatology monthly wave height and wind speed as the input of the DNN. However, there is still no improvement.

The authors seem to be interested to develop a wind measurement system that competes with a satellite scatterometer or altimeter, but this project is inherently dealing with in an in situ platform. Is not the goal to develop an in situ system that has precision and accuracy metrics similar to those of the 10 min averaged wind anemometers used at sea?

Although the data is obtained from an in situ platform, the rationale of this model is more similar to satellite scatterometer and altimeter that use the surface wave properties to indirectly retrieve the wind. The author thinks that indirectly estimated wind should be compared with the indirectly estimated wind. Meanwhile, both scatterometer and altimeter are regarded as successful remote sensors for wind speed retrieval, especially the scatterometer. As an indirect estimation model, being comparable with a scatterometer indicates that this model is already practical for many applications such as model assimilation. That is why the author mentioned remote sensing in the text several times.

Of course, it will be nice if the precision of the model can be similar to the anemometers. However, it is difficult for a model trained against the anemometer data (which is regarded as the "ground truth") to reach the same accuracy. Another problem is that if there is no better "ground truth", it seems also to be difficult to judge whether the accuracy of an indirect wind-estimation model is better or worse or similar than the anemometer data.