Comments on AC1:

Reply to Reviewer #1:

Indeed, as described in the preprint that the sea surface wind and waves are important parameters for the marine environment and ocean dynamics. This also implies that the interactions between them involve complex dynamic procedures resulting in the intricacies of coupling between them that make their individual characteristics difficult to resolve. Buoys on one hand though with limited amounts and distributions have been long providing good measurements of both wind and wave parameters respectively and simultaneously, on the other hand making complementary to remotely sensed wind and waves from satellites. The wind-wave interaction can then be modelled from buoy observations, while deep learning provides powerful tools in non-linear modelling and regression. The author thereby applies a deep neural network (DNN) for extracting wind information from wave spectrums provided by buoys for further applications to buoys without wind measuring ability benefitting such buoys with lower costs. The motivation and origin of this research are reasonable and good.

First, the author would like to thank the reviewer for the positive opinion on the motivation of this work and the comments which are helpful for the improvement of the manuscript. Some revisions are made to the manuscript according to them. For some comments the author has different opinions on, explanations are given in this reply. The author hopes the reviewer can change his opinion and find the merit of this manuscript.

Of course, the interaction between wind and wave is very complicated, and the aim of this manuscript is not to reveal how waves (or to say, wind-sea) grow under the force of wind. The aim of the work is simply to establish a practical method of estimating wind speed and direction from wave spectrum measurements. The result of the model also indicates that this model to estimate sea surface wind from wave information can be useful. Therefore, the author submitted this work to *AMT* which is a journal about technology instead of a more "physical oceanography" journal such as JPO and JGR-Ocean.

Review comments to AC1 (ReAC1)_1: As stated in the previous comments, the validity of applications of AI in ocean science is challenging and triggers the recent highlighting of causality considered in the procedures of the model establishment. This is utterly lacking for this research, and not corrected from the replies. Besides, it is not likely that the result of this model described being useful. Moreover, AMT is about technology, correct applications of reasonable science are the fundamental of good technology, the arguments on science basis form the basis or the important part of the theories in applied technologies. Redo thoroughly this research is still necessary for many aspects. Specific comments are as following:

Unfortunately, the research in this preprint falls into the trap set by that the DNN theory that can fit all models provided wide enough (though which is true mathematically). This may be due to ignoring that the meaning of the DNN model expressed is data and inputs-outputs dependent or self-consistent within such boundary. The model is only physically meaningful than mathematical results when not only the data or inputs are of good quality but also considering underlying physical principles to an extent a DNN can resolve. This can also be expressed with the state for one of the challenges for the application of artificial intelligence in ocean science: moving from purely statistical prediction to process-based models that embody causal relationships (Catalán, I., A. Solana, et al, 2021).

The author does not think this work "falls into the trap" of DNN theory. Firstly, a model needs not to have explicit physical meaning to be useful. Nonparametric empirical models and methods are widely used in many aspects of ocean science. For example, the operational algorithms of many ocean remote sensors (e.g., the sea-state bias correction for altimetry, the D-Matrix algorithm for microwave radiometer, the watercolor algorithm for type II water, to name but a few) are also data-based and empirical. These models made many contributions to the development of ocean sciences and technologies. Secondly, the model presented in this work is not without physical bases or causal relationships. The work is based on the simplest idea that there is some quantitive causal relation between local wind and waves because waves are generated by wind. Although the explicit form of the wind-wave relationship is unknown, and the DNN can "learn" such a relationship using a large amount of data. In the author's opinion, artificial intelligence is the most suitable for such regression problems in ocean science: we know there are some causal relationships between inputs and outputs, but the physical model is too complicated to establish.

ReAC1_2: The problem of the proposed model lies in that the "some" causal relationships are without specific research but thrown into the DNN tools, which can result easily in applications in the unknown region the model cannot be representative. While the training sets cannot cover all different combinations of wind speed, wind direction, wave height/slope, wave direction, and the environmental parameter affecting different relations between them. Usually, in empirical models, such problems are seriously treated by specific analysis of the features and to how much extent the inputs and outputs are related, and what inputs cannot be applied. For example, empirical sea-state bias correction for altimetry is generally based on models of specific air-sea interaction as well as surface scattering methods of electromagnetic waves. As for 'D--Matrix' approach which seeks a linear relationship between measured SSM/I brightness temperatures and environment parameters, it is rather complex and uses matrix coefficients based n particular seasons and latitude bands that the measurements were taken from, and has a root in the measuring principles of radiometers. While there's no proper reason that the applications of the more powerful AI methods for regression excel requirements in this point. At the same time, the author argues about "a model needs not to have explicit physical meaning to be useful", and uses observed directly and unexplained in all aspects of spatiotemporal and statistical features of the inputs and outputs.

More specifically for this research, it applies the spectrum parameters all at once as inputs for wind speed and direction ignoring the underlying multi-scale heterogeneity in time and space due to the complex relation of the interface interaction that can be embodied by a spectrum interpreting them in different approximations of the governing equation for energy distributed for different k values. Such approximation cannot only be expressed in another way round by expanding the observed energy distributed for different k values (spectrums are fitting of the observations) for another fitting from DNN. In other words, here DNN makes little extra contribution than the observed spectrum from this research. What is captured by DNN cannot be clarified makes things worse. Around Line 115, from the results, "the wave spectrum might also better reflect wind information a short period before" is contradictory to the fact that wind-wave spectrum ranges from lower frequencies to higher frequencies due to momentum transformation between waves of different lengths. For wind estimation, short wave measurement is relevant while the modulations from longer waves are non-negligible, from tilting effects to the short waves or modification of amplitudes of the short waves by exchange of energy altering atmosphere conditions close to the sea surface.

The author does not understand why "DNN makes little extra contribution than the observed spectrum from this research". The DNN here is simply a model to estimate wind information from the input wave spectrum. This task might be done without the DNN, but no other models can perform as well as DNN at this stage, as far as the author knows (Please let the author know if there is a better model). Also, the author is confused about "spectrums are fitting of the observations". The observation of what? According to the understanding of the author, the wave spectra from the buoy can already be regarded as the observations of waves instead of a fitting.

The author believes that the expression, "the wave spectra might also better reflect wind information a short period before", is not contradictory to the fact that wind-wave spectrum ranges from lower frequencies to higher frequencies due to momentum transformation between waves of different lengths. But the author agrees that the expression is not precise and might cause some misunderstanding. Here, the "wave spectrum" only means the wind-sea spectrum measured by buoys, which needs some time to respond to the wind force (under the action of wind input, dissipation, quadruplet wave-wave interaction, etc.). Since the wind-sea spectrum is more impacted by the wind information a short period before than by the current wind, it is OK to say that the buoy wind-sea spectrum can better reflect the wind information a short period before than the current wind. The sentence has been revised to "the wave spectra of gravity waves from buoys might also better reflect the wind information a short period before than the current wind information", which should be more accurate. Our results showed that the DNN perform better when using the wave spectra to estimate the wind information one hour ago than to estimate the current wind, which also supports the above opinion. Of course, the "short period" here is not a specific time and can be from several minutes to several hours. Here we use the wave spectra to estimate the wind information one hour ago simply because the temporal resolution of the data is one hour.

In fact, the main reason for the author to directly apply the spectrum parameters all at once as inputs of the DNN is precisely to take into account the modulations of longer swells on shorter wind seas (another reason, which is not that important, is to illustrate how easily such a practical wind-estimation model can be established). If the modulation of swells is important for the estimation of wind information from buoy wave spectra, their impacts will be easily "learned" by the DNN (as demonstrated in many studies of wind remote sensing, e.g, Stopa et al. 2016 [Scatterometer], Li et al. 2018 [SAR], Jiang et al. 2020 [Altimeter]), and the low-frequency part will be important inputs of the wind-estimation model. However, the sensitivity test in Section 4 shows that the spectral information at frequencies lower than 0.1 Hz (mainly swells) does not have a significant impact on the model output, which indicates such modulations are not crucial for the estimation of wind from the wave spectrum. We have now pointed it out explicitly in the text that "Previous studies of wind remote sensing showed that the modulation of swells on capillary waves has some impacts on the wind speed retrievals (e.g., Stopa et al. 2016, Li et al., 2018, Jiang et al. 2020). However, according to the results here, the swell's modulation on wind-seas has little impact on wind-estimation using buoy wave spectra."

ReAC1_3: Again, the problem does not lie in if there is a more powerful model than DNN for complex problems (despite there being many other AI methods suitable for even chaotic situations), but the way in which this research is modeling. In addition to the previous comments, even if now it is clarified that the wave spectrum specifically refers to as the wind-sea partition the problem still exists. Here is a more detailed explanation as it seems a bit brief in the previous version.

For most of the NDBC buoys, the directly measured parameters are not the spectrum parameters. Note the ocean waves in different lengths embracing each other in a complicated way that is apparently non-linear and is still without a final answer due to unresolved air-sea interaction for wavelength ranging in a range of spatial scales. Hence the transfer from measurements of buoys to spectrum in different frequencies include basic assumptions on their interactions (the hourly wave height measurements from NDBC are not enough for an exact wave spectrum). And different empirical spectrums can be classified into this category. Although there is some new type of NDBC buoys that measure spectrums directly (the α parameter, et al.), and the amount is about 100 around half-half near-shore and off-shore. The samples are far from enough to cover all value space of different combinations of what also mentioned above as effecting factors for relating winds from waves: the interaction of air to sea, and the energy transfer as well as respond interactions between waves of different lengths: the near-surface air condition, wind speeds, wind directions, wave speeds, wave directions, et al., to form a steady model for such a complex problem facing only slightly better situations modeling wind and wind-induced waves in issues for calculating spectrum in the buoys when they cannot be measured. Besides, the off-shore and near-shore regions are with different features, and the locations are also limiting the conditions of sampling, in addition to the fatal lack of analysis of data inputs and outputs as well as related analysis (the comparison with the remotely sensed wind will be discussed in the next comment), how it can provide predictions from limited samples are not obtained from the established DNN model. Then this research is making no extra contribution than the spectrum coefficients from limited sampling. Though by applying parameters mimic to [1] helps narrow down the uncertainty space, while the lack of sampling can cause problems. And the conclusion that “the swell’s modulation on wind-seas has little impact on wind estimation using buoy wave spectra” may also be due to the defect of the model established, while in [1], this is also considered in the parameter β.

[1] Voermans, J. J., Smit, P. B., Janssen, T. T., and Babanin, A. V.: Estimating Wind Speed and Direction Using Wave Spectra, Journal of Geophysical Research: Oceans, 125, 10.1029/2019jc015717, 2020.

Moreover, though the training procedure is mathematically accomplishable, as in the preprint, where the results can be validated in error analysis from the testing set. Let alone the comparison of results to remotely sensed winds are not validated ignoring representative features of remote sensing results and buoy observations. Buoys generally provide the spot-based measurement of winds while remote sensing results are averages of a large region. The distributions of samples for each wind (and direction) bin are not discussed, the sample number may be skewed due to distributions of nature winds, while such effects are ignored in this research.

The comparison with buoy-measured wind is almost a common practice in the validation of wind products from different types of remote sensors (e.g., scatterometer, altimeter, SAR). The community understands there is a representativeness error between remote sensing results and buoy observations (also between different remote sensors because they cannot measure exactly the same region at exactly the same time). The author even has a paper focus on mitigating the impact of this issue (Jiang 2020). However, buoy and remote sensing data are still comparable because of the potential equivalence between (remote sensing) spatial and (in-situ) temporal average, and also because most geophysical parameters do not vary severely in a small spatial-temporal domain. Otherwise, the comparison between any data from different types of data sources (in-situ, remote sensing, numerical model, etc.) will be problematic, which is not helpful for the development of science and technology.

The wind direction is relatively uniform as seen in Figure 2d/2f. The condition that wind speeds are not uniformly distributed can lead to the results that the model performs the best near the peak of the probability distribution. However, the empirical probability distribution function has been shown in Figure 2. The results indicate that the trained DNN model performs not that well for extreme winds (e.g., RMSE > 3 m/s for U10 > 20 m/s), due to insufficient numbers of samples in high wind speed. This is not surprising as the air-sea interaction becomes much more complicated during extreme wind (e.g., spray). The errors of estimated wind speed and direction as the function of measured wind speed have been shown in Figure 2c and 2f, which gives more details of the model's error for different wind speeds, which is also a guidance to the user of the model. We can see from Figure 2c that the model has the best performance for 3~10 m/s wind speed, and the RSME remains lower than 2 m/s for 1~17 m/s wind speed. This indicate the skewed sample numbers does not have large impacts on the model in moderate wind conditions.

ReAC1_4: The previous comment on this issue was brief and the point was not made clear, sorry about that! The comparisons of remotely sensed winds and buoy winds are typical and useful. However, in the research, the buoy wind is applied directly for matching with the scatterometer products, and reasons are as provided before, buoy winds are instant measurements while the remotely sensed winds are spatially averaged, both values cannot be compared directly, pre-processing are required. Meanwhile, this differs from the SSH measurements of buoys in that they can be compared with their nearest spatiotemporal remotely sensed match, for the SSH recorded are time-averaged values that are somehow equivalent with spatially averaged measurement. This is also the case for the wave period parameters.

Although some sensitive analysis for inputs as the selection of frequency discussed in part 4, this was unfortunately misinterpreted as well, due to the little effort taken for understanding the relation between observed inputs and outputs. This is similar to the results part around line 115, longer waves are with wind information that cannot be resolved by the mapping to winds from the DNN established directly fitting the observations.

As mentioned in the above response, the information of long waves is used as the inputs of DNN. If the reviewer is familiar with machine learning, he/she will know that the impact of swell on wind estimation can be easily captured by a DNN. The author is not saying that longer waves (frequency < 0.1 Hz) are without wind information, but the DNN results tell us the spectra of low-frequency waves provide no additional help for the estimation of wind speed and direction.

ReAC1_5: See ReAC1_3.

The discussions following such content are not proper as well. When the model boundary is not clear due to the aspects listed above, there is little chance for these DNN models to apply in QC procedures or other applications. The results are also not likely to be improved including compact wave drifters, as the air-sea interaction in different scales is not likely to be well described in the reasons above.

The author is not sure about what the reviewer means about the model boundary. But the author does not want to argue too much on whether this DNN model can be applied in QC procedures. The data in Figure 3 has already proved that some bad-quality data is identified using the DNN. This is the best evidence to show that the usefulness of this model in the QC of buoy data. It is noted that these bad-quality data were not identified in the QC procedures of the National Data Buoy Center.

The author needs to emphasize that the function of this DNN model is to estimate wind information from wave spectra instead of gives a better explanation about the physics of air-sea interaction. Regarding whether the results can be improved including compact wave drifters, the manuscript has shown that the high-frequency information is crucial for buoy-wave-spectrum-based wind estimation and the accuracy of the model is impacted by the ocean current. The data from such drifters can contain better-quality wave spectra (due to their better response to short waves) with more high-frequency information and also the ocean current information. For DNNs, better and more relevant inputs can usually give better output. That is why the author believes the results can be improved.

ReAC1_6: In addition to previous comments, the QC of a model not well established doesn't help.

To wrap up, for such a model without awareness of the causalities between the inputs and outputs, especially under the circumstances such causalities are complex and wraps between even inputs and outputs, the deductions made based on them can easily go wrong. This is exactly the case for wind-wave interactions, such that improper analysis generally appears here and there for this preprint.

As mentioned in a previous response, the causality between input and output is considered, but not in any explicit form. The author does not deny that this model can go wrong sometimes, especially in very low and extreme wind speeds. However, this has been discussed in the manuscript and the error functions of wind speed and direction were given. As a model to estimate wind speed and direction, there is no need to judge right or wrong, there is only accurate and inaccurate. As a famous saying goes, "all models are wrong, but some are useful". This model provides an accuracy of ~1.1 m/s for wind speed and ~14° for wave direction, which should be regarded as useful.

ReAC1_7: See previous comments. Besides, the saying is a warning to not stray into the mistake of choosing one's model as correct over reality. This is consistent with ReAC1_1 somehow.

2) Around line 30, as mentioned before, short gravity-capillary waves are modulated by longer waves, though in the case of scatterometry, the orbital velocity of longer waves cannot be observed, and the tilting effect may not be obvious for them modulated to gather on the crests, by modulating the surface wind stress that changes the amplitudes of the short waves, which cannot be ignored, the long wave information does exist in scatterometer observations.

Many studies have shown that the modulation of longer waves can impact the wind estimation of scatterometers. But it seems to be difficult to retrieve wave information directly using these impacts. To the best of the author's knowledge, there is no effective model that can obtain wave information from the scatterometer backscatter data independently. Therefore, it is OK to say wave information is not available from scatterometers.

ReAC1_8: In fact, they are available in scatterometer observations, for example:

a) Wright, J.: Backscattering from capillary waves with application to sea clutter, IEEE Transactions on Antennas and Propagation, 14, 749-754, 10.1109/tap.1966.1138799, 1966.
b) Plant, W. J.: in: Surface Waves and Fluxes, Springer, Dordrecht, https://doi.org/10.1007/978-94-009-0627-3_2, 1990.
c) Quilfen, Y., Chapron, B., Collard, F., and Vandemark, D.: Relationship between ERS Scatterometer Measurement and Integrated Wind and Wave Parameters, Journal of Atmospheric and Oceanic Technology, 21, 368-373, 10.1175/1520-0426(2004)021<0368:rbesma>2.0.co;2, 2004.

4) Around line 35, low temporal resolutions do not cause low performance near shore. There are near-shore products from scatterometers for example. Besides, inter-constellation will solve the coverage problems to an extent.

The author simply wants to state: 1) space-borne remote sensors often have limited temporal resolutions, 2) space-borne remote sensors often perform badly near shore. That is why an "and" is used instead of "so that". To make this point clearer and accurate, this sentence is revised to "…space-borne remote sensors often have limited temporal resolutions and they often perform worse in nearshore regions…".

Of course, more satellites can increase the temporal resolution and spatial coverage. This is common sense that is not related to the theme of this manuscript so that the author thinks there is no need to mention it.

ReAC1_9: "1) space-borne remote sensors often have limited temporal resolutions," is also common knowledge should not mention if "more satellites can increase the temporal resolution and spatial coverage." is not needed.

2) Near shore are not necessarily poor or worse in performance, there are already examples many years ago:

a) Chelton, D. B., Schlax, M. G., Freilich, M. H. & Milliff, R. F. (2004). Satellite Measurements Reveal Persistent Small-Scale Features in Ocean Winds. Science, 303, 978--983. doi: 10.1126/science.1091901

b) Chelton, D. B., Freilich, M. H., Sienkiewicz, J. M., & Von Ahn, J. M. (2006). On the Use of QuikSCAT Scatterometer Measurements of Surface Winds for Marine Weather Prediction, Monthly Weather Review, 134(8), 2055-2071

5) Around line 45, again, direct comparisons of buoy results are remote sensing products are not validated.

This point has been explained in the previous response. The comparison between buoy and remote sensing results is a common practice. If this is not valid, the comparison of almost any data from two different sources will be invalid.

ReAC1_10: It is not validated for this research since the comparisons are not properly done, but not due to the theory to make comparisons between buoys and RS results are invalid. See also ReAC1_4.