Reply to Reviewer #1 :

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.

The author would like to thank the reviewer again for his helpful discussion. The author still thinks that AI is best suited for the problem that we know there are some causal relationships between inputs and outputs, but the physical model is too sophisticated to establish. For a problem, if the full causality and the underlying physics are well understood and parameterized, there is no need to introduce AI anymore. Therefore, for such a deep learning-based empirical model, whether such a model is useful or not should be judged based on the performance of the model instead of its underlying physics. Also, the author needs to point out that this is not a model without physics concerns. The physics background of estimating wind information from wave spectra has been discussed in Voermans et al. (2020) so that this work focus on the wind-estimation model itself.

Voermans J. J., Smit, P. B., Janssen, T., and Babanin, A. V.: Estimating wind speed and direction using wave spectra. J. Geophys. Res. Ocean, 125, 2019JC015717, doi:10.1029/2019JC015717, 2020.

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.

As mentioned before, the physics background of estimating wind information from wave spectra has been discussed in Voermans et al. (2020). Their results and analysis have shown that the wave spectra can be used as the input of an empirical model for wind estimation. However, there might be more factors that might be difficult to take into full account in the semi-analytical method (e.g., the swell modulation effects mentioned by the reviewer, although this study shows it is not very important). That is why the author used a DNN model and also why the DNN can give a robust estimation of the wind information. If only some irrelevant parameters are used to train the DNN, the DNN will not give a good result. In other words, if there is no reasonable physics sustaining this model, this model cannot work that well.

Regarding the two examples, sea-state bias correction and D-Matrix, the author did not mean that they are not without physical bases. In these models, the physics background is known to some extent but not completely known. There are also some factors that might be difficult to take into full account in the analytical or semi-analytical models. Therefore, after the theoretical studies have shown what can be used as the inputs of a model, scientists established effective empirical models to bypass the complicity of some high-order processes. For instance, after the theoretical analysis shows that the sea-state bias can be linked to the wave height (Hs) and backscatter cross-section ($\sigma 0$), a data-based empirical Hs-$\sigma 0$ look-up table can be used to estimate the sea-state bias.

In fact, the logic of establishing this DNN model is very similar to that of establishing the D-Matrix algorithm. Regarding the causal relationship between input and output in D-Matrix, we know that the change of geophysical parameters, such as SST, water vapor, and wind, will impact the received radiance of different channels. However, the analytical form of the relationship between them is difficult to know. Similarly, for this problem, we know that the wind will impact the buoy-measured wave spectrum, but the analytical form of the relationship is also difficult to know. At that age, the training of a DNN is much more difficult to train and the concept of AI is not that popular, thus, scientists assume the relationship to be linear and use in-situ observations to train the linear D-Matrix. Similarly, this study uses the DNN and a large amount of data to find the relationship between inputs and outputs. We know today that it is also OK (and even better) to use a DNN to establish the relation between geophysical parameters and the brightness temperature of different channels of radiometers.

The author is also aware that the DNN can be over-fitted and can be inapplicable for the condition that is not covered in the training dataset. That is exactly the reason why the data used in this study were divided into the training set and validation set, each containing 50% of the data. The evaluation of the model's performance was only conducted in the independent validation set. According to the concern of the reviewer regarding whether the model is applicable for different locations, the author also tried to divide the training set and validation set according to the buoys' location. We use the data from buoys 45001-51101 (53 buoys) as the training set and the buoys 41002-44066 (48 buoys) as the validation set. The locations, wind-wave climate, and other environmental properties are significantly different for the two sets (none of the buoys in the validation set is in the same basin of the data of the training set). However, the results remain quite good as shown in the figure below (Figure R1). This again shows that the resulting model can adapt the condition for different regions, showing the robustness of the model.

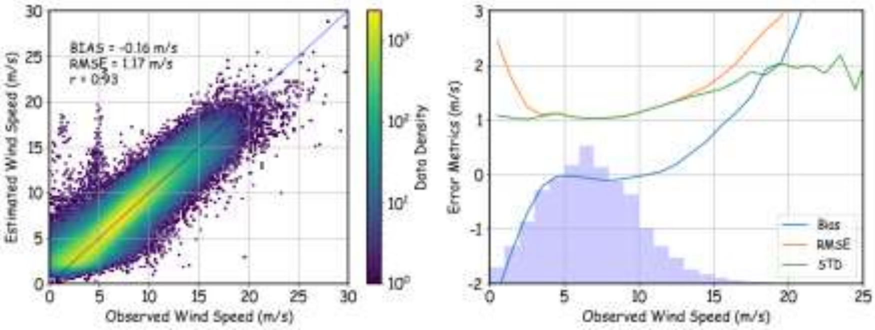

Figure R1. *The same as Figure 2b and 2c in the manuscript, but the training and validation sets are different. In this figure, the data from buoys 45001-51101 are used as the training set and the buoys 41002-44066 are used as the validation set, and the comparison is only conducted in the validation set.*

Regarding what should not be used as the model input, it is one of the advantages of the DNN model. If one input term is not important for the output, the DNN can "automatically" ignore the impact of it given the sample number is large. For instance, we can add more irrelevant inputs for the wind-estimation DNN model, which will not impact the result as the weight of these inputs will be set to zero during the training process of the DNN.

ReAC1_2: 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.

"A model needs not to have explicit physical meaning to be useful" is just the general attitude of the author. The author wants to express that even if a model is not explicitly physical meaningful, it can still be useful sometimes. For example, the AI models of pattern and speech recognition still seem to be far away from the explicit physical meaning. However, these models are already widely used in many aspects.

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.*

The wave spectrum parameters from NDBC buoys used in this study are all derived from the translational or pitch-roll information from the accelerometers and inclinometers onboard buoys. A fast Fourier transform is applied to the sensors' time series (~20 minutes) to transform the data from the temporal domain into the frequency domain. Therefore, the buoys used in this study are all new types of buoys referred to by the reviewer, and all the spectra used in this study are directly measured instead of fitted. Of course, the measurement of the spectrum (using any method) is based on the assumption of the quasi-stationary random process and weak nonlinearity, which is why the sampling time for wave measurement can neither be too long or too short. It is noted that the concept of the wave spectrum (and the Fourier theory) itself is based on the assumption of linear superposition of waves with different scales.

Although the reviewer thinks that the samples of these buoys might not be enough to cover all value space of different combinations of effecting factors, the sample size of more than 1 million records is not a small one. Besides, many values in the effecting factor space without samples can still be obtained using the interpolation and extrapolation ability of the DNN. This can also be illustrated using the result in Figure R1. The samples from the training set should not be able to cover all combinations of factors in the validation set, as they are data from different basins. However, the model still performs well in the validation set.

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.

ReAC1_3: 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*

Although the off-shore and near-shore regions are with different features, these features do not necessarily impact the estimation of wind from wave spectra. Figure 1 in the manuscript has shown that the accuracy of wind in off-shore and near-shore regions are not significantly different. If whether a buoy is off-shore or near-shore is significant for the wind estimation, the performance of the DNN will be improved by including this factor. The author tried to include the distance to coast (and also tried the condition near/off-shore using the 50km offshore criterion) as the input of the DNN, and the model did not give a better output. Similarly, if the swell modulation is important for such a wind-estimation model, the model's residuals should be significantly correlated to the swell information and such correlation can be easily identified by a DNN and included in the model.

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.

First, this research did not match the buoy wind with the scatterometer products. The collocation is only made between buoy wind and buoy spectra. Second, the wind measured by buoys are also time average so that can be compared with remote sensing (spatially averaged) wind product. As mentioned in the previous reply, such comparison is common practice in the validation of wind products of remote sensors including wind products from scatterometers and altimeters.

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

Figure 3 in the manuscript (Figure 4 in the revised manuscript) has already proved that some bad-quality data is identified. These data were not identified in the NDBC QC procedure.

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.

This has been also mentioned in the reply to ReAC1_1. The physics background of estimating wind information from wave spectra has been discussed in Voermans et al. (2020) so that this work focus the wind-estimation model itself.

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)0212.0.co;2, 2004.

The "wave information" in the context refers to information of gravity wave (instead of capillary wave) such as wave spectrum, wave height, wave period, or wave direction. These wave parameters are not available from the scatterometer data product. In the three reference provided by the reviewer, the wave parameters are also unavailable from scatterometers.

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

According to the suggestion of the reviewer, this sentence has been revised to "space-borne remote sensors often perform worse in nearshore regions than in the open ocean due to the land contamination of backscatter" (the expression about the limited temporal resolution is removed from this sentence). Yes, the performance of remote sensors are not necessarily worse near shore, but the retrievals are often (not always) impacted by land contamination of backscatter near shore.

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.

This has been discussed in the reply to ReAC1_4. This study itself did not involve remote sensing data at all. The RMSEs between remote sensing and in-situ wind (~1 m/s and 15 °) are from many previous papers and is only used as an error reference for the model in this study.

The author thanks the reviewer again for these comments.