Response to reviewers

21st of June 2020

1 General comment to reviewer # 1

We would like to thank reviewer #1 for his comment on the manuscript with the working title of "Norwegian hindcast archive (NORA3) - A validation of offshore wind resources in the North Sea and Norwegian Sea". Please see the response below.

1.1 Response to comments from reviewer #1

RC1: The authors do not convincingly show what is novel and/or more advanced in their model compared to others. E.g., from Table 1 I cannot see what is new/better compared to NEWA. The authors discuss that NEWA is not validated offshore; however, this is different from having something in the model which is novel/advanced compared to other models. If the authors could show that NORA3 fits measured data significantly better than competing models/datasets, it could be considered an advancement; however, this is not done.

AC1: Thank you for pointing this out. The novelty of NORA3 over other similar data sets is that NORA3 is created by another numerical weather prediction model (in this case the HARMONIE-AROME, Cy 40h1.2, which have spectral dynamics which is very different from the WRF model which solves the momentum equation on a staggered grid). Currently all available wind resource data sets are created by the Weather Reasearch and Forecasting model (WRF). The creation of NORA3 by a different NWP model will contribute to a diversity in the available wind resource data sets. When a multi-model ensemble of these data sets are considered for wind power planning the wind resource uncertainty can be quantified and reduced in regions like offshore which often have very few and short time in situ observational references. We know from inter-comparison of numerical weather prediction models, that there is no such ting as a overall "best" model. Quality depends strongly on which spatial and temporal scales are of interest, geography, the selected measure of quality and the application the data will be used for. Thus, the usefulness of multi model ensembles has become increasingly clear over the last few decades in research field such as weather prediction and climate change. We also see this slowly trickling down into wind resource research community. In addition, this study contains novel validation measures not seen in the assessment of other wind resource data sets covering the North Sea, Norwegian Sea and Barents Sea. Finally, we tend to agree that papers focusing on verification of new simulations seldom are extremely novel. However, they fill a very important role by documenting quality. This is a prerequisite to ensure a sound and critical use of the simulations by stakeholders and fellow researchers.

NORA3 are extensively validated towards observations and compared against the host data set in the paper of Haakenstad et al. (2021). Nevertheless, we have now added to the paper a validation and comparison of NORA3 and ERA5 towards observations. A quantile-quantile plot (qqplot) for each of the six sites is generated and will be added to the paper to compare the wind speed distribution between the three data sets (obs, NORA3 and ERA5). The qqplots show that NORA3 performs better than ERA5 for almost all wind speed percentiles, especially for the high wind speeds. In addition, a seasonal validation is performed, showing the mean values and standard deviations (std) for the three data sets for the four seasons (DJF, MAM, JJA, SON). The seasonal analysis show that both mean and std values from NORA3 are closer to the observed mean and std values for all six sites. Two tables showing the seasonal results will be added to the paper.

As several new wind resource data set are becoming available a model comparison is indeed of great interest and something we will pursue in future work. However, given the page limitations and the fact that this is the first paper to evaluate the wind resource estimates from this simulation set we have prioritized to put the focus on a detailed comparison against observations instead of introducing a second objective to the paper.

The following section will be added to the paper (added text in italic font):

Comparison of NORA3 and ERA-5

The NORA3 wind estimates in 10 masl are extensively validated against observations and compared against the ERA-5 reanalysis in [1]. Nevertheless, we compare the performance of NORA3 and ERA5 towards the observed wind speed climatology in the six wind sensor heights (68-140 masl).

The observed seasonal average and standard deviation of the wind speed are shown in Table 3 and Table 4, respectively. In addition, the relative difference between the observations and NORA3 (n3 (%)) and the observations and ERA5 (e5 (%)) are also shown. Table 3 shows that the modeled average seasonal wind speed from NORA3 are consistently closer to the observed values for all the seasons. The standard deviation (std) is here a measure of the variability in the wind speed. The seasonal variability is shown in Table 4. Compared to ERA5, NORA3 is consistently closer to the observed seasonal std for all the six sites.

Figure 3 shows the quantile-quantile plot (qq-plot) between the observed wind speed and modeled wind speed by NORA3 and ERA5. The qq-plot determines if the two data sets are drawn from the same sample distribution. If the circles lie on the reference line the data sets comes from the same data distribution. For all the six sites the models perform best for the lowest wind speeds ($u \leq 10 ms^{-1}$). For both models the deviation from the reference line ("ref line") increases with *increasing wind speed percentile. Nevertheless, NORA3 is consistently closer to the reference line compared to ERA5, and especially for wind speed exceeding a typical cut-off wind speed. A technical feature called "high wind ride through" enables the turbine to exploit more of the very strong wind speeds ($u \geq u_{co}$). In offshore areas higher winds is happening more frequently. Therefor, the importance for a NWP model to estimate these strong wind events correctly increases. NORA3 outperforms ERA5 for these high wind speeds ($u \geq u_{co}$).*

As illustrated in Fig 3 the largest difference between the observations, NORA3 and ERA-5 is found for wind speeds exceeding a typical cut-out limit of 25 ms^{-1} ($u \geq u_{co}$). Since the power production is terminated or at least reduced when $u \geq u_{co}$ we calculate the wind power capacity factors (CF) for the three data sets. This is done to see how the the models perform in terms of power production where the strongest wind speeds not influence the result due the production cut-out limit. Table 4 contains the CF for the observed data, NORA3, and ERA-5 for the six sites. NORA3 performs consistently better than ERA5. For the six sites used in this study, NORA3 is 1.8 percentage point closer to the observed average CF-value compared to ERA-5.

The validation of wind climatology in NORA3 and ERA5 show that the downscaling of ERA5 in the process of creating NORA3 has resulted in an improved wind resource data set. The remainder of this study will focus on the validation of NORA3 towards observed wind climatology.

RC2: The statements such as "…NORA3 data is rather well suited…", "…slightly conservative…" and "…slightly underestimated…" are very vague. What are they based on? E.g., "The model is relatively good…": relative to what?

AC2: Thanks for pointing this out. We believe that you found these statements in the Abstract where we write "NORA3 data is rather well suited for wind power estimates, but gives slightly conservative estimates on the offshore wind metrics". The next sentences in the abstract explain what we mean by these phrases: "Wind speeds are typically 5 % (0.5 m/s) lower than observed wind speeds, giving an underestimation of offshore wind power of 10 %-20 % (equivalent to an underestimation of 3 percentage point in the capacity factor),for a selected turbine type and hub height. The model is biased towards lower wind power estimates because of overestimation of the frequency of low-speed wind events ($< 10 m/s$) and underestimation of high-speed wind events ($> 10 m/s$). The hourly wind speed and wind power variability are slightly underestimated in NORA3. However, the number of hours with zero power production (around 12 % of the time) is fairly well captured, while the duration of each of these events is slightly overestimated,leading to 25-year return values for zero-power duration being too high for four of the six sites. The model is relatively good at capturing spatial co-variability in hourly wind power production among the sites. However, the observed de-correlation length was estimated to be 426 km, whereas the model-based length was 16 % longer".

However, we will add some explanatory text where we have written these vague statements without further explanation.

RC3: The biggest issue I have with the presented results is that they are not compared to other models. I thus consider it impossible to judge whether the model results are good or bad (see also the previous comment related to this). As NEWA mesoscale data are available (https://map.neweuropeanwindatlas.eu/), they should be compared to the NORA3 model results. I also consider comparison to ERA5 to be required, as the downscaled data should be shown to outperform the data they are based on.

AC3: Thanks for your comment. Please see AC1. I understand that Table 1 and section 2.1.1 creates an expectations towards a comparison of NORA3 and the other data sets listed in the table. Therefor, Table 1 and section 2.1.1 will be removed from the revised manuscript to make the intention of this study more clear, namely a thorough validation of NORA3 towards offshore observations. The introduction will be rewritten accordingly to clarify the intention of the study.

RC4: As a large portion of the paper is comparing wind speed distributions, I suggest the authors to compare the distribution results also to microscale resolution data, which are available both from NEWA and GWA (https://globalwindatlas.info/).

AC4: Thank you for the suggestion of comparing NORA3 with microscale resolution data. This is of great interest since comparing NORA3 with microscale resolution data will determine how well NORA3 is capturing meso- and small scale features. However, as mention in AC1 we have prioritized to focus on a detailed comparison against observations instead of introducing several new aspects to the paper. Hence, comparing to microscale resolution data will be further work.

RC5: The authors write: "...underestimation of 3 percentage point in the capacity factor". This would seem to me a large error, e.g., when considering profitability of an offshore wind power plant, and does not seem to be in line with the other statements of the model being quite good (see the comment 2) above). Please comment.

AC5: This is an interesting point raised by the reviewer and we tend to agree. It highlights how rather small wind errors translates into sizable wind power errors. Since a cost-function like the levelized cost of energy(LCOE) is inverse proportional to the wind power produced during the lifetime of the wind farm, using NORA3 as wind power production estimator will result in a wind farm profitability being on average 6% higher than first expected. Given that the required rate of return when planning offshore wind projects typically is 5-10 %, a deficiency of 3 percentage point in the CF is an sizable error. This highlight the need for building up archives of simulations like the NORA3, to be able to conduct informed uncertainty calculations for the production in regions where observations are not existing or are very sparse.

However, a comparison between NORA3 and ERA5 in terms of the wind power capacity factor (CF) shows that the ERA5-based CFs are on average 5 percentage point (approximately 10% difference in CF) lower than the observational based CFs. Hence, the improvements using NORA3 over ERA5 gives more realistic wind power profitability measures. I will add a comment to the paper regarding the profitability, and the difference between obs, NORA3 and ERA5.

RC6: I do not understand what this sentence means: "The corresponding wind speed variability is given by the Weibull standard deviation (std). The Weibull std is used instead of the Gaussian std because of the shape of the wind speed distribution (see Fig. 2a for an example wind speed distribution)." Standard deviation (or variance) gives the 2nd moment information of any (well-defined continuous) distribution (https://en.wikipedia.org/wiki/Standard_deviation); it is not related to the shape of the distribution. I find the sentence very confusing, please clarify and give references to what a "Weibull std" is and why it should be used.

AC6: We are sorry for the confusion regarding the choice of standard deviation (std). What was meant was that the std was not calculated directly from the observations, but by first fitting a Weibull distribution to the data before the std was calculated. The idea was that this would be somewhat more robust then calculating it directly because the extreme tail of the empirical distribution was reduced with the Weibull fit and therefor made a smaller impact on the std. We have now removed this and use a traditional std calculate directly from the empirical data, since the qualitative information is the same and it is easier to understand.

RC7: If you consider standard deviation to be an inadequate measure, why not use higher moment measures (such as kurtosis, skewness) and/or compare the entire PDFs/CDFs of the distributions?

AC7: Figure 2 is showing the wind distribution and differences in the simulated distribution, compared to the observed. In the discussion of the figures we discuss the quality of the higher moments in the simulation.

RC8: Why only 2004 to 2018 are simulated? ERA5 data are available for tens of years; why not downscale more years? Would not using longer time series provide a better estimate especially of events occurring very rarely? (the authors focus on rare events in specific sections)

AC8: Thanks for referring to the somewhat short model period of NORA3 compared to ERA5. We have only validated 2004-2016 (study period changed after exclusion of FINO1 from 2010-2018) due to the currently available data from NORA3. NORA3 is continuously being generated, and at the time of writing it covers 1994-2020. When the model integration is finalized (in autumn 2021) the NORA3 data will cover the time period from 1979 to present, and will be regularly updated in the coming years when ERA5 data becomes available.

RC9: The authors write: "Hence, the observed wind speed is somewhat more intermittent and variable than the modeled wind speed, indicating that HARMONIEAROME is missing some of the high-frequency variability embedded in the wind field." Why do you say that especially high-frequency variability is missing? Variance (and thus standard deviation) is a sum of all spectral components (https://en.wikipedia.org/wiki/Spectral_density) and thus a mismatch in standard deviations could be caused by mismatch in the spectra at any frequencies.

AC9: Thank you so much for noticing this badly explained phrase. I agree that standard deviation and variance encloses all temporal scales. During the generation process of the paper we performed a spectral filtering of the observational and modeled data using a Butterworth filter. This result was not implemented as a part of the paper. Performing this data-filtering illustrated that NORA3 was missing out some high-frequency variability. I will add an explanatory text to the paper to clarify.

RC10: I find it difficult to understand what exactly is presented and done in sections 2.6, 4.4 and 4.5. I understand that for the estimation of extreme (very high) wind speeds, extreme value theory may be needed to extrapolate beyond the limited time range of data (however, please note comment 8) above related to this). But I find it very confusing that two very different causes for "zero wind power events" are mixed: one being quite frequent (wind speed below cut-in speed), which does not seem to need extreme value theory for estimating the likelihood; and one being relatively rare (wind speed above cut-out speed). But even the latter should happen offshore many times during a multi-year dataset. Why does it need extreme value theory? I would consider extreme value theory to be needed to estimate values like 50-year maximum wind speed, which may be needed for the design of the turbine (but that is very different from wind speeds causing storm shutdowns).

AC10: Thanks for your comment of the extreme value analysis on zero power events. I agree that performing an extreme value analysis on the number of zero events will not work since the the occurrence of these events are not a rare event. However, we have performed an extreme value analyses on the maximum duration of a zero-event and not the occurrence of the events. We calculate the maximum duration (in hours) of a zero event expected to happen at least once during the lifetime of a wind turbine (25 years) due to either too low ($u < u_c i$) or too high ($u \geq u_{co}$) wind speeds. Knowing about the expected extreme duration of zero-event is useful information for planning the amount of regulating power needed in a grid to ensure the needed electricity supply at all times, as well as for planned turbine maintenance and for the profitability of the wind park. Therefor, we have validated NORA3's ability to reproduce the observational based maximum duration of a zero power event.

RC11: I find sections 4.4. and 4.5 confusing. a) Why do you apply extreme value theory? (see the previous comment); b) How is the extreme value theory exactly used to give you estimate of how long the "zero-events" last? Why does this matter, e.g., for LCOE? (I think LCOE is impacted by the entire generation distribution, not just the "zero-event" likelihood). I do not understand what the last paragraph on 4.5 is saying, consider a thorough rewrite.

AC11: Thank you for this comment. Please see AC10. I agree that the LCOE is determined from the entire generation distribution. However, the generation distribution also includes the part of the distribution caused by either too low or too high wind to produce wind power, hence the zero power generation. Since, the overarching goal of an operating wind farm is to produce as much as possible, it is crucial to know about how frequent a zero power event is, and also the maximum expected length of these events during the turbine lifetime due to scheduled maintenance and as information on how reliable the production from the wind farm is when large amounts of wind energy is planned to be integrated into the grid. The extreme value theory is conducted in the following way:

- Firstly, the Kolmogorov-Smirnov p-test is performed to test the null-hypothesis. The null-hypothesis states: the data is **not** drawn from the chosen data distribution (GEV or Pareto). Testing the null-hypothesis is done by calculating the distance between the empirical (either "block maxima" (BM) or "peak over threshold" (POT)) and theoretical (extreme value (GEV) and generalized Pareto distribution (Pareto)) cumulative distribution functions.

- Secondly, since the result from the Kolmogorov-Smirnov test tell us that we cannot exclude the possibility that the data is drawn from either of the two data distributions we fit the observational based and model based extreme zero-event durations to GEV and Pareto and find the corresponding 25-year return values for the following five sites: Ekofisk, Sleipner, Gullfaks C, Draugen and Heidrun. Fino1 is excluded from the extreme value analysis due to the shorter time series (2004-2009).

I will add explanatory text to clarify the sections encompassing the extreme value theory and analysis.

References

[1] H. Haakenstad, Ø. Breivik, B. R. Furevik, M. Reistad, P. Bohlinger, and O. J. Aarsnes. NORA3: A non-hydrostatic high-resolution hindcast for the North Sea, the Norwegian Sea and the Barents Sea. *Submitted to Journal of Applied Meteorology and Climatology*, 2021.