

We thank both the reviewers for their enthusiasm about this research and for conducting a thorough review to improve our manuscript, we greatly appreciate it. We have provided responses to each of the specific comments by reviewer below, highlighting the modifications we have made to the manuscript.

## Reviewer 1 Comments

This is a very relevant and important topic, which has analysed in great detail here, resulting in insights and recommendations for further improving the accuracy of real-world wind resource assessments. In general, the paper could benefit from some graphical overviews of the applied analysis methods, because it is very difficult to lose track of exactly what has been done and what has been compared to what - especially given the large number of acronyms used. I have made specific comments directly in the paper in the attached document.

**Author:** We thank the reviewer for their interest in this research and for conducting a thorough review to improve our manuscript, we greatly appreciate it. We have provided responses to each of the specific comments by reviewer below, highlighting the modifications we have made to the manuscript.

**Reviewer** (line 1): Usually written in present tense

**Author:** We use the present tense to describe the field and past tense to present the work that we have done in the past as we have done throughout the paper, barring a few exceptions the reviewer has alerted us to in later comments.

**Reviewer** (line 5): It's not clear exactly what you mean here. OK it's confusing because 6.5 is less than 6.9, so one would usually write "6.5% to 6.9%). But it's the other way round because it's negative in the introduction. I suggest changing them to 6.5%-6.9% and 5.5%-6.7%.

**Author:** We rephrased this sentence to clarify our point that our study yielded similar results, but with a slightly increasing overestimation bias, and include the negative signs to reiterate the point and offer further clarity.

**Reviewer** (line 12): This section could benefit from more details about the importance of quantifying the uncertainties of EYAs, as well as a part describing the sources of uncertainties.

**Author:** We have expanded the first introduction paragraph to speak to more of the EYA generation process and its importance for project financing. We have also followed this paragraph with an explanation of EYA and operational analysis PXX values.

**Reviewer** (line 15-16): This sounds strange without referring to the specific case. For how many wind farms was this analysis done, how large were they, where were they located,

how much data was there, etc.?

**Author:** This section was broken out into two additional paragraphs (combined with some of the reviewed second paragraph) to go into more details of each study in a more cohesive manner.

**Reviewer** (line 16-17): I don't understand what this means. Please explain better.

**Author:** We have rephrased the latter half of the sentence to specify an additional 1-2% of the bias could be stemming from unreported losses that the study's authors were unable to account for.

**Reviewer** (line 18-19): Again, this needs to be put into context by describing the details of this study.

**Author:** As mentioned, we have now broken out the study comparison portion of the paragraph into two separate paragraphs with this study getting its own attention.

**Reviewer** (line 19): Do you have a reference for this, or how do you know?

**Author:** We both added the unreferenced standard deviation value and the missing reference, which was the same as the one in the previous sentence.

**Reviewer** (line 30): I suggest present tense

**Author:** Thank you for flagging, we have now adopted the present tense for this phrasing.

**Reviewer** (line 33): makes

**Author:** Thank you for flagging, this is now fixed.

**Reviewer** (line 35): delete one "in"

**Author:** Fixed, thank you for flagging.

**Reviewer** (line 38-41): I would write this in present tense

**Author:** I have updated the verb tenses throughout the paragraph, and made minor revisions to the phrasing to accommodate the shift in verb tense.

**Reviewer** (line 43-44): Explain how P90 is calculated from P50 and the uncertainty.

**Author:** I have added an explanation that in the EYA process, the P50, P90, and uncertainty are derived from a Gaussian distribution that is largely mirrored from the updated introduction.

**Reviewer** (line 44): This is a strange sentence for two reasons: (1) It refers indirectly to P50 and P90 from the previous sentence. Why not directly just refer to what you introduced in the previous sentence? Not every reader will understand what you are referring to here. (2) You say that you also considered uncertainty, but I think you need the uncertainty to calculate P90 from the P50, don't you? Please reword.

**Author:** In conjunction with the above change, we have clarified the differences in EYA an

Monte Carlo (our study) processes for understanding the P50, P90, and uncertainty values as its own paragraph before going into further discussions of our methods.

**Reviewer** (line 52-53): You are mixing the passive and active voice in one sentence - I would avoid this.

**Author:** The mixing is of which verb is active vs passive has been flipped so that the sentence reads primarily in an active voice.

**Reviewer** (line 53): is the word "data" maybe missing after "utilized"?

**Author:** The phrasing has been revised to no longer contain a missing word and clarify that we worked with many of the same industry partners to obtain data.

**Reviewer** (line 55): I'm getting confused about the method and where different types of data are coming from. I suggest a graphic to clarify it.

**Author:** The previous sentence now clarifies that the industry-provided MORs contained the monthly energy production and losses. The breadth of these MORs are simply being described in this sentence.

**Reviewer** (line 59): I'm not sure why this information is presented in box plots. Could you explain the relevance? Why is it important to visualise the distribution of these factors?

**Author:** The boxplots are demonstrating the types of wind plants that were analyzed in this study. The decision to use a swarm plot overlaid with box and whiskers chart as opposed to other chart types, such as a histogram, was mostly driven by the desire for a compact, yet legible visualization of the data's distributions. In particular, the box plot can identify where the data are concentrated, and the statistical distribution while the individual swarm points can highlight where clusters exist within the data. The juxtaposition of these two elements helps highlight that we are working with a diverse spread of wind plants for this project. We prioritized the compactness as this is not the most impactful visual in the narrative, but aids in understanding the kinds of plants we analyzed. We have also clarified this sentence to ensure readers understand that we are merely demonstrating that there is a wide variety of plants represented in this study.

**Reviewer** (line 60): Meaning that there were no MORs? Or what's the connection between this statement and the fact that there are fewer results for some variables?

**Author:** We expanded this sentence and the prior one to further clarify that we are discussing the project variables collected from the EYAs that are associated with the monthly operating data.

**Reviewer** (line 72): Is there a reference or guideline on this? I.e. which correlation coefficient should be reached in order for the reanalysis conditions to be "highly suitable"?

**Author:** We've removed "highly," but there is not a general recommendation other than a

basic correlation analysis of data with less than a moderately correlated relationship would be unsuitable. We did check on some values, and found those with very weak or negative correlations were consistently outliers in our results, and have included this in our improved discussion of the correlation analysis.

**Reviewer** (Figure 2): If you are going to make the effort to show this distribution of correlation coefficients, I would suggest a short discussion of reasons for deviation. Does the correlation coefficient reduce with distance from the data point, for example?

**Author:** We have added a second paragraph for the discussion of reanalysis correlation that describes some of the potential factors for why the reanalysis data may be unsuitable for the removed plants. The primary reasons include complex terrain, potentially unaccounted for wakes from upstream wind farms, and reanalysis data being unable to fully account for ground-level wind conditions in complex terrain due to both the nature of atmospheric modeling and the use of a coarse grid.

**Reviewer** (line 85): Is this always available on the turbine level? What do you do if not, remove the month?

**Author:** We have clarified in the second sentence of the *Wind Plants* subsection that our data are monthly plant-level aggregations, and not turbine-level data. There were also no instances of missing data between the starting and ending period of the provided operational data, so an explanation of this methodology was not provided.

**Reviewer** (line 101): The method described below would be easier to understand with a flow diagram

**Author:** We were unable to coerce the operational analysis considerations into a flow diagram, but have instead converted this comma separated listing of parameters and the following bulleted list into a single table. This table shows both which variables are varied between individual simulations and which are varied between analyses.

**Reviewer** (line 102): Why?

**Author:** We have indicated that this number is the model's default, and added a clarifying sentence about the model defaults in general. We have also noted that the default value choices were not specifically documented, only the choices for the values relative to each other.

**Reviewer** (line 103): Why?

**Author:** We have indicated that this number is the model's default, and added a clarifying sentence about the model defaults in general. We have also noted that the default value choices were not specifically documented, only the choices for the values relative to each other.

**Reviewer** (line 104): What is this?

**Author:** We have provided a definition of the windiness correction and reference to the paper where this term is first introduced.

**Reviewer** (line 106): Reference?

**Author:** Thank you for flagging the missing reference, we have now added it.

**Reviewer** (line 130-133): This seems misplaced - is it supposed to be summarising the results shown below? If so, it is usual to put this after presenting the result, not before. If not, then it should refer to a plot or table and then describe this fully.

**Author:** We have replaced this paragraph with a short description of the two results sections.

**Reviewer** (line 130): reported? expected?

**Author:** As stated in the previous response, we have removed this paragraph entirely in place of a more general description of the section.

**Reviewer** (figure 5): I would add an analysis of the distribution, as well as just a comparison of the averages. Otherwise why show this plot?

**Author:** We have added more discussion about the distribution of the results.

**Reviewer** (line 161): Can these be compared directly here?

**Author:** Mostly yes. Lunacek, et al., 2018 forms the basis of this body of work, and the methods themselves directly influence those that this study took and what OpenOA's Monte Carlo AEP methods are (alongside Bodini, et al., 2016). However, they are also not directly comparable because they include different sets of wind plants. But they both include a large number of plants with regional diversity and are expected to represent industry-wide trends.

**Reviewer** (figure 7): Can you say something about the statistical significance of the differences between biases vs. consultant (and also vs. year from Figure 6)? Given the large range of biases, it seems that the difference between the averages presented here may not be statistically significant. Can you do this using the uncertainties in the next figures using a t-test?

**Author:** This is a good point, we have now performed and provided the results of a one-way ANOVA to demonstrate there is no significant difference between consultancies with respect to their P50 predictions.

**Reviewer** (figure 7): Can you discuss other factors that may affect this variation, such as site location, turbine type, terrain complexity, etc.?

**Author:** We have added a further discussion of other factors, namely COD, OEM and geographic regions. For the regions, we were unable to acquire permissions from our

partners prior to publishing our final response to include a similar plot as the consultants. For factors such as terrain complexity or induced wakes from plants built upstream, we decided that there isn't enough time to fully quantify these relationships.

**Reviewer** (line 219): Quantify this

**Author:** We have added the range of differences for the most and least inclusive subset results.

## Reviewer 2 Comments

This is a high-quality and highly valuable paper that makes an important contribution to the wind energy community. In particular, it will help investors and financiers gain greater confidence in the assumptions underlying their investment decisions, and ultimately reducing risk and encouraging long-term investment in wind power projects.

I've included several comments in the attached pdf, focusing mainly on how the results could be presented in a way that is more accessible to industry professionals. My goal is to be able to share direct excerpts from the paper with colleagues who may not have a technical background, using language and framing that make the topic and its relevance more immediately clear.

I've also included some technical questions and suggestions for areas that could be explored further. However, these represent potential avenues for future work, and the current paper already feels complete and well-developed as it stands.

Thank you for addressing such an important topic!

**Author:** We thank the reviewer for their enthusiasm about this research and for conducting a thorough review to improve our manuscript, we greatly appreciate it. We have provided responses to each of the specific comments by reviewer below, highlighting the modifications we have made to the manuscript.

**Reviewer** (title): typo

**Author:** Thank you for flagging this, it now reads as "annual."

**Reviewer** (line 4): Didn't prior studies suggest a narrowing of the bias?

**Author:** We have rephrased the beginning of the introduction to dedicate a paragraph to each of the original study, and our former colleagues' study. The former is directly comparable, though the latter paints a better picture, but is not comparable as it is a meta-analysis of mostly results that have not gone through peer review.

**Reviewer** (line 4-8): These are really important results! If investors and banks are going to finance projects, then reducing the risk is key to achieving this long-term, and production deltas and uncertainties are a key risk!

**Author:** Thank you, we are glad that the results will be having a real work impact.

**Reviewer** (line 36): strikethrough

**Author:** Fixed, thank you for flagging.

**Reviewer** (line 65): Is this resolution correct for MERRA2? Normally its 0.5 ° x 0.625 °, but perhaps this is a downscaled dataset?

**Author:** This is correct and have amended the text, thank you for flagging.

**Reviewer** (line 104-106): Often there are months that are "outliers" when it comes to losses, e.g. a line outage, grid outage, or curtailments, but is it fair to treat these as outliers and normalise the data, given they often reoccur?

**Author:** While it isn't completely fair to remove all the months as they can reoccur, they are still not representative of long-term trends, and tend to occur earlier in the plant's lifetime. Ultimately, this does not change the conclusions of the paper, which is that there is a significant EYA overprediction. We have added additional commentary in the text to describe this justification.

**Reviewer** (figure 4): Why is there any spread at all? In my view there shouldn't be any doubt about the Net Energy (i.e. revenue data) each month, so slightly odd that they'd be reported differently between EIA and MORs.

**Author:** We have clarified that there is a lack of specification about what data are provided by plant owners, and any potential processing steps before being made available in the API. There is a general (though undocumented) concern with EIA data that annual numbers are provided and then disaggregated to monthly production using regional trends, and so we find it necessary to demonstrate that there is good cohesion between EIA- and owner-provided energy generation figures.

**Reviewer** (line 146): This is a huge delta. Have you got an idea as to why the EIA delta is much larger than the MOR data too?

Also, is it possible to state what the current "P50" in pre-construction estimates is (e.g. its really a P25)?

Or conversely, can we say what P-value I should assume to get to a true P50 from an pre-con EYA?

We're always trying to work out if a "hair-cut" should just be applied to the pre-con EYA results we receive.

**Author:** Similar to the response to Reviewer 1, there isn't any documentation about how the EIA receives or handles the reported net energy data, so it's difficult to pinpoint any particular factor that could be causing the 3 percentage point bias discrepancy. One thing worth noting is that there is significantly more EIA data available, but it's unclear how this would make impact given the high correlation evinced in Figure 4.

We have also added a discussion of the true P-value for EYA P50s and the long-term corrected P50's P-value relative to the EYA following the prediction error results.

**Reviewer** (figure 5): One thing that everyone wants to avoid are the projects with significant shortfalls, i.e. outside the normal distribution. Are there any findings to suggest which sites these are, and what the key failing is?

**Author:** We have added a further discussion of other factors, namely COD, OEM and geographic regions following the discussion of Figure 6. For factors such as terrain complexity or induced wakes from plants built upstream, but had decided that there isn't enough time to fully quantify these relationships.

**Reviewer** (line 165): Was this due to confidentiality, or are these results from internal analyses? For example by utilities rather than consultancies?

**Author:** I updated the phrasing to read "no consultant name being provided" to clarify that we simply do not have the information available to us to group them into a specific consultant because in many cases the data in the EYA was provided to us by the owners or operators, rather than as a full report or by the consultants. This latter clarification is also now addressed earlier in the manuscript. It should be noted that any EYA providing entity is being referred to as a consultant, which they would be for the purposes of the EYA.

**Reviewer** (line 170): I've seen a tendency in EYAs to apply unknowns as uncertainties only, rather than to reduce the P50 and then apply an even distribution for the uncertainty. Is this evidence of that?

**Author:** We don't have enough data to support this since most of our EYA data collection only consists of the primary energy generation data with some project metadata compared to a complete EYA report with documented assumptions. What kind of data do you believe would support this finding?

**Reviewer** (figure 9): An interesting result. So a 1-year P90 from a pre-construction EYA is conservative?

Financing is typically done on a 10-year P-value, are those results the same?

**Author:** Yes, but not overwhelmingly so. We discuss later that the 10<sup>th</sup> percentile of errors is still negative, rejecting this as a universal finding. It is also worth noting that the figure compares the long-term corrected P90 and the EYA P90, which are not directly comparable as the long-term corrected uncertainty is generally much lower, and is why we discuss the prediction errors in Figure 13. We have also swapped this figure out with the long-term P90 bias, which exhibits a similar trend.

**Reviewer** (line 193-198): Is the issue that everyone is assuming a normal distribution, but the tail is much longer than expected?

**Author:** If the P50 was generally accurate, but the P90 was significantly overpredicted, then we could probably conclude that a normal distribution doesn't accurately reflect the uncertainty. However, our results show that the P50 results are overpredicted, and would skew the P90 results as an artifact.

**Reviewer** (line 225): HVS1?

**Author:** This has been replaced with a reference to Lunacek et al., 2018. Internally, that study was phase 1 of the historical validation survey (HVS) project, hence HVS1, and the paper was drafted using that shorthand. Thank you for flagging this lingering reference.