

## ***Interactive comment on “The Making of the New European Wind Atlas, Part 1: Model Sensitivity” by Andrea N. Hahmann et al.***

**Anonymous Referee #2**

Received and published: 17 April 2020

### General Comments

The paper summarizes an exhaustive sensitivity analysis performed to inform the final model setup of the New European Wind Atlas. This surely must be the most extensive such analysis to date and overall is an impressive achievement. The novel use of the Earth Mover’s Distance is also applauded and clearly offers a much-needed complementary metric alongside the typical timeseries-based performance metrics.

I believe this paper should ultimately be published; however, I have several comments and concerns about the work that have not been addressed in the paper. First, all of the critical validation was performed in Northern Europe, despite the NEWA being produced for Europe and Turkey as a whole. I realize that computational expense and data availability/quality were probably a factor, I can’t help but feel that with such collab-

C1

oration across European institutes that a more regionally diverse validation campaign could have been performed. Of course NEWA has already been produced, but I think some critical commentary on how validation in Northern Europe (with its unique climatology) would apply across other climates in Europe with their own unique climatologies is needed here. Otherwise, the paper reads as if the idea of more extensive validation was overlooked.

Furthermore, I did not find sufficient presentation of results to justify selection of the final model setup. Rather, a wind profile plot and two heat maps of bias and EMD were provided, and it seemed very quickly the section was wrapped up with the final model selection. I think some further synthesis is required, such as a table of figure showing mean bias, RMSE, EMD, etc. across all validation sites. Without this, in my opinion, the selection of the final model setup seems unjustified.

Finally, as far as I can tell, ERA-interim was used in the sensitivity analysis, but ERA-5 was used in the final production run. This point is not discussed in this paper but I think it’s an important one. Does existing research suggest bias or EMD differences between the two data sets? If so, what are the implications on selecting the best model setup using one large-scale forcing but pivoting to a new product for the actual production runs?

In conclusion, I think this is a valuable contribution to the literature. However, several key limitations of this study need to be sufficiently addressed and discussed before final publication. In addition, a couple summary figures and tables would help justify final model selection.

### Specific Comments

Page 1, Line 9: Why were sensitivity experiments only conducted in Northern Europe when the data set was for Europe as a whole? Surely tall masts must be available elsewhere? If this was a decision based on computational restrictions, this should be stated and the implications of this smaller validation domain, in the context of regional

C2

wind climates, should be discussed.

Page 2, Line 15: Can 'linearized model' be described more, or at least a couple references listed to provide background?

Figure 1: As in comment in Line 9, validation only in Northern Europe poses a problem for a product that covers Europe as a whole. This key study limitation needs to be discussed in detail.

Table 1: What is the time resolution of the observed data used to indicate sample size? I'd assume hourly but please make this clear.

Page 6, Line 9: Given the known impact of turbine wakes at these measurement sites, why not filter the data by wind direction to ensure the data are free stream? Especially in such a detailed sensitivity analysis where performance metrics between different model setups can be on the order of 0.1 m/s, allowing wakes to affect the measurement data seems inappropriate.

Page 7, Line 14: I'd use 'interpreted' rather than 'understood' when describing EMD as a measure of physical work.

Page 7, Line 15: Given the novelty of the EMD metric, I wonder if a new Figure showing the area between cumulative distribution functions would be useful, given this is how the metric is actually computed.

Page 7, Line 16: What are circular variables and why are they relevant here? Are you validating wind direction?

Page 8, Line 23: Why was WRF 3.6.1 used, given it is 6 years old and the significant advances made since then? Was this part of an older study that is now being published?

Page 10, Line 3: But MYNN winds are higher in the NW offshore domain and lower in the SW domain. Can you discuss? Is NW offshore domain generally more stable?

C3

Figure 6: Given the detailed justification of EMD earlier, why is it not being used here?

Figure 8: I'm struggling trying to distinguish the different model runs. Multiple setups seem to have identical markers (at least to the naked eye). Also the lines are so tightly clustered that it's generally not possible to discern one profile from another. As such the Figure does not provide much useful information and I would recommend revising or deleting.

Figure 9a: Would an additional column showing average across sites be useful in identifying the best performing model setup?

Figure 9b: I'm not sure I see the value of performance metrics relative to the 'base' setup. In my mind this base setup is just another member of the ensemble and not otherwise special. So why compare all ensembles against this one? Do we know it to be the most accurate? If not, I don't see the value in this relative comparison. Please justify.

Figure 10b: Likewise to comment above. I'm not seeing the value of this relative comparison.

Page 18, Line 5: This is a big jump to conclude the best performing model setup based on the figures shown in this section. For example, the improved performance of MO over the Base and MM5 setups isn't clear from the profile plots or the heat maps. I think some final figure or table is needed showing key performance metrics averaged across all sites in order to justify this model choice.

It also seems that the multi-physics sensitivity analyses and the selection of final production run in Section 5.3 was done using ERA-interim as the large scale forcing in WRF. However, ERA-5 was used in the final NEWA. This seems problematic given potential differences (e.g., biases) between the two data sets. I understand that ERA-5 was not available at the time these simulations were performed; however, some discussion around the implications of changing the large scale forcing without sensitivity

C4

analysis needs to be provided.

Page 19, Line 8: Unclear how ERA5 reanalysis slow down of winds relates to a sensitivity analysis of ERA-interim, FNL, and MERRA2. Was ERA5 part of this comparison?

Figure 11a and 12a: What is the difference between BASE and ERAI? I thought the base run was done using ERA-interim.

Figure 11b and 12b: Same comment as previous.

---

Interactive comment on Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2019-349>, 2020.