# Response to Referee #1

(With minor changes from the version available online).

Thank you for the comprehensive comments, and also for taking the time to truly read through our manuscript. We feel that your comments were very helpful for increasing the quality of the paper to its current level. Your comments, together with those of referee #2, led to a thorough revision of the paper.

The most general comments regarding the revisions to the manuscript are:

1. At the start of the research project typically there are high expectations placed on the sensitivity experiments, however, reality always brings some corrections and caveats. Given the enormous possibilities in setting up WRF, an "optimal" configuration is unreachable. We have tried to revise the introduction to convey that the paper focuses on finding the "best possible" model configuration **constrained** by the practical issues in running the model simulations and the ultimate goal to use the simulations for a **wind atlas**.

2. We humbly think that the present analysis offers and unprecedented assessment in terms of the variety of WRF model configurations tested.

3. The manuscript aims to tell the story of how the NEWA wind atlas came to be. Therefore, further analysis of the model results will make the flow of the paper less clear. We have tried to enhance this structure in the revised manuscript.

4. We have replaced some of the figures (6, 9–12, 13) to homogenise the analysis of the results. We have also added new figures including the RMSE and circular EMD for wind direction.

5. We strengthened the connection to the companion paper, https://www.geosci-model-dev-discuss.net/gmd-2020-23/, which is now available.

The reviewers' comments are in black and our responses in blue.

## General comments

1. First and foremost, the dataset could (should?) be made publicly available. The data availability section only refers to the final NEWA product, not to the sensitivity experiments upon which the presented results have been based. This is not just a reproducibility issue. Many interesting research and model development questions beyond the scope of NEWA can be addressed with this rich sensitivity dataset, and it would be a waste not to share it.

   We agree. However, the subset of the WRF model data from the simulations totals 15 TB. We are trying to find a solution, perhaps via an EUDAT grant. As a minimum we will make available the yearly wind statistics from each simulation in Zenodo.

2. Furthermore, the discussion is very limited in scope. There is no comparison with similar efforts (although smaller in scope), based on different models. The discussion stays away from any physical interpretation and lacks critical reflection on important choices that have been made. The impact beyond the NEWA project is not considered at all. For example, the authors state that "it would have been optimal to evaluate the results of the ensemble simulations with the large dataset used in the companion paper". But this can still be done, and although the insights would not propagate to NEWA, they could clarify some of the questions that currently remain unanswered. The last of the specific comments lists further issues that I would like to see in the discussion.

   We agree with the statements above, however we feel that the narrow nature of the manuscript is justified taking into account that there exists a second part to this study in a companion manuscript that is dedicated to critically evaluating the results of the final choice. Also, as mentioned above, the manuscript tries to explain the rationale of the choices made during the creation of NEWA wind atlas. Including further data analysis at this stage is beyond the scope of this objective and may mislead the reader in the purpose of the manuscript. We believe the scope of the manuscript is still useful to modellers both in wind energy and in wider applications of numerical weather prediction models.

   Also, we believe the manuscript is already long enough. Even if we wished to do further evaluation with the data used in part 2, it would not be possible. Figure 4 in part 2 shows the tall masts available in the Vestas database. The comparison would have been limited since there are no masts in Denmark and very few in Germany. In Poland data exists in some masts but for only a few months in 2015.

3. Some minor aspects of the model configurations are not documented, which hampers reproducibility. For example the determination of vertical levels or the parameters of the lambert projection. Perhaps the authors could share the namelist of the final configuration? It is also not clear whether the WaSP downscaling methods has been applied to the presented results (and if so, it should be documented).

   All namelists are shared in the project GitHub (https://github.com/newa-wind) and in the NEWA Zenodo (http://doi.org/10.5281/zenodo.3709088 site, this also includes all the "geo" files used in the model simulations. The link to zenodo is located under "assets" in the manuscript GMD website. Therefore, the simulations are reproducible. The full description of the NEWA model grid is in the companion manuscript [1].

   As for the second question. No further downscaling is done in this Part 1. All the evaluations against tall masts are done with the raw WRF model data. The sites are relatively simple (offshore and over flat terrain), where the microscale modelling will add little to the mesoscale model solution.

## Specific issues

1. P2 L14: While it is very clear in the abstract, I miss a sentence like: "This paper describes our efforts to find an optimal configuration of the WRF mesoscale weather model for the production of a New European Wind Atlas (NEWA)." in the introduction. The configuration of WRF for the production of NEWA is the main focus of the paper, yet its introduction is a bit out of the blue with a reference to Petersen 2017. It would be good to provide more context about NEWA. Why was WRF chosen, for example? Given that virtually all options within WRF are investigated in this study, presenting the choice for WRF itself as a an accomplished fact feels a bit unsatisfactory. Line 14 in particular starts with "Given the EWA is 30 years old", which begs for something like "A and B bundled forces to produce an updated wind atlas."

   Agreed. We have added a new paragraph about the wider NEWA project in the introduction. The rationale for using the WRF model is also included.

2. P2 L10: perhaps explain "the so-called wind atlas method" in one or two sentences? Is this the same method referred to in P2 L22? And is this method also used for the evaluations presented here? P23 L31 makes me think it is indeed, yet P24 L17 seems to suggest the opposite (but it is a bit unclear what is meant by "the full downscaling model chain"). If no further downscaling is used for this study, perhaps don't mention it at all.

   The wind atlas method is mentioned in the introduction because it was used to create the earlier European Wind Atlas. But we understand that can be confusing and reference to it was removed from the rest of the paper.

3. P3 L 12-29: this paragraphs seems a bit out of place. I suggest moving it to somewhere around P3 L4-6, such that P3 L30 logically follows after the part about "The approach in NEWA". Perhaps the statement about "best practice setup" can then also be combined with the reference setup referred to in P3 L34.

   Excellent suggestion, thanks. The three paragraphs starting in P2, L25 and ending in P3, L29 have now been restructured in a more logical way: (1) adapting models to wind energy applications, (2) review of previous ensemble studies, and (3) approach taken in this paper.

4. P4 L20: This requires further discussion, as land surface/soil moisture "memory" is known to significantly affect the results.

   We did consider the issue of land surface and soil moisture memory. In regional climate model simulations, the NWP model parameterisation are often tuned to avoid model drift (e.g. [3]). For generating a wind atlas, we are not worried about model drift because the simulations are re-initialised often (here - every 7 days). It would be optimal if we could re-initialise the atmospheric model often, but keep the state of the land surface from one simulation to the next. However, for practical reasons this is not possible, since the simulations had to be run sequentially. Also, it is not obvious that the precipitation produced by the WRF model is accurate enough to keep the soil moisture from drifting. Lastly, the land surface and soil moisture memory is indeed important in simulating climate-relevant parameters such as temperature and precipitation, but we don't think there is enough evidence that it is critical in wind reanalysis. The connection between soil moisture, sensible heat flux, wind profile and wind speed does exist, but we do not have a systematic way of validating it in the context of NEWA. We have tried simulations initialised with Global Land Data Assimilation System (GLDAS) data, but the results were inconclusive.

A reference and edits have been made to the text in this part.

5. P5 Fig1: All masts seem to be located in the northernmost domains (compare with Fig3). If the configuration was optimized for Northern Europe, what does this mean for the validity of NEWA for the South-European domains?

   Yes, unfortunately there are very few quality tall (above 50 m height) masts in Europe publicly available. At the start of the project, we had eight sites in Northern Europe and another hand-full in the other domains. It is not optimal, but we had no other option. Winds observations from surface stations are plentiful, but often they are placed in complex sites that make the evaluation difficult, and the accuracy of a model at 10-m height does not give any warranty that the WRF configuration will also be accurate at wind turbine height [2]. In this manuscript we argue that the combinations of PBL/SL parameterisation schemes behave the same way in different regions in Europe (section 5.1) and thus conclusions from Northern Europe are also applicable to the other regions. We believe that this statement and approach is supported by the evaluation results of the wind atlas that are described in companion manuscript (Part 2)

6. P6 L8: "Due to difficulty . . . has not been filtered or corrected". This requires more justification. At least the authors could say something about how the performance differs between the various masts or between wind direction sectors. That should provide some intuition about the potential effect of wind farm distortions. It might also be relevant to mention the wind directions that were filtered for FINO, Riso and Hovsore explicitly. Are these prevailing wind directions or not? And how do they relate to the nearby coastlines? Especially in coastal areas, I think it is not safe to assume that model performance is uniform across all wind directions.

   The wind sector filtered for mast distortion is now explicitly listed in Table 1. We have also corrected the height of FINO1 and FINO2 and added the height of the wind direction data used in the analysis.

7. P6 L15: While I believe the presented evaluation metrics achieve the stated objective of selecting a single best model configuration for the production of the NEWA

(in terms of wind speed), their presentation is quite unclear. I would advise to use the more common term "mean absolute error" (MAE) instead of "absolute bias". Also I would advise against making all metrics "relative", which is mostly confusing. Comparison against a baseline (or: reference) is very good. However, isn't the more common approach to use their fraction rather than the difference? See for example literature on fractional skill score, or the excellent textbook by Wilks (statistical methods in the atmospheric sciences). You would get SS = 1 - (MAE / MAE_ref), and SS = 1 - (EMD / EMD_ref), which would approach 1 for a perfect forecast and 0 for no improvement over the baseline. I suppose that such a uniform scoring system would help to judge whether an improvement in one metric is worthwhile if it is accompanied by deteriorating scores for other metrics or locations. Right now, that"s not clear (see e.g. my specific comment P18 L5).

That is a very good suggestion, thanks. We have now revised Figures 9–12 to use this "skill score (SS)" as $SS = 1. - (MM/MM_{ref})$, where $MM$ is BIAS, RMSE, EMD or CEMD.

8. P6 L19: "The main goal of the NEWA project was the evaluation of the wind climate, which is usually understood as the probability distribution of wind speed and direction at a specific point". Why then, is wind direction not evaluated at all in this manuscript? And what about vertical wind shear?

   That is a very good point. We have now included the evaluation of the wind direction for the initial simulations and the large ensemble. However, as mentioned above, the manuscript tells a story of how we arrived to the final NEWA configuration. Adding new parameters such as wind shear, while interesting and relevant, deviate from the main story of the document.

9. P6 L24: This statement is quite irrelevant and I doubt if it's always true. I suggest to remove it.

   Agreed. The two sentences have been removed.

10. P6 L29: Move part about RMSE to after the stuff about bias. Also, perhaps refer to a paper about skill-scores. Part about comparing to baseline/reference setup is a good idea and might be useful for others that want to learn from this study. Therefore, a very clear explanation is appropriate. I had to read it three times.

    Agreed. The section has been rewritten to objectively describe of the methods used. We also added the new metric for the wind direction, the circular EMD, or CEMD.

11. P7 Fig2: I understand that the histogram representation of the wind speed distribution is appealing because it is widely known. Panel A succeeds in showing the

difference between EMD and absolute bias, but I wonder if this cumulative distribution plot would be even more intuitive. Also, I'm curious why the difference between EMD and absolute bias is larger for small absolute bias.

An additional panel showing the EMD as the area between the cumulative distributions has been added to Figure 2.

The question "why the difference between EMD and absolute bias is larger for small absolute bias" could be reformulated as "why is the difference between EMD and absolute bias smaller for larger absolute bias". If two distributions have the same mean, then the bias is not able to distinguish between them but the EMD can be used to measure how similar are the distributions. If two distributions have the same shape but different means, then the minimal transport necessary to "move" the distributions towards each other will be equivalent to the difference in means. A nice illustrative example of EMD properties can be found in: Lupu et al.[6]. We have added this reference to the article's text.

12. P7 L15: The EMD explained as the are between CDFs is very intuitive. It took some effort to verify this, but eventually I found it (https://stats.stackexchange.com/a/299391). It seems that this statement is only true for univariate distributions. A reference here would be appropriate.

    Agreed. The reference Rabin et al. [7] extends the CDF interpretation of EMD to circular variables. The text has been updated to clarify this and the fact that this applies only to one-dimensional distributions.

13. P8 L16: I understand that the authors try to put emphasis on the differences (or rather: the absence thereof) between the geographical domains, especially seeing that PBL is further investigated later on. It is indeed a good idea to test this domain-sensitivity with various set-ups. But the section is written such, that the reader tends to focus mostly on the performance between PBL schemes rather than geographical domain. This is especially true towards the end of the section, where it seems that conclusions are drawn about the reference configuration, rather than about the domains. Both figures 5 and 6 contribute to this shift of focus.

    Agreed. The main focus of the section was to show similar sensitivity in various regions and not on the evaluation. We suggest a new structure where section 5.1 relates to the five domains only, and a new section (5.2) is added where we discuss the validation against the sites in only one of these domains.

14. P8 L19: I think it would be good to briefly explain the differences between these two PBL schemes, and why these two schemes were chosen. PS: or in the later section.

    Good point. A sentence has been added to the revised manuscript.

15. P9 Fig3: The experimental sites don't seem to correspond to the locations of the masts used for the evaluation presented in this paper. What then, is the reason to show these sites? Perhaps this figure could be merged with Figure 2? Also, the abbreviation "PD" is not clear to me.

    We removed the NEWA experimental sites from the figure. They are not relevant to this paper because data from the experiments were not used in the model validation. "PD" stands for Perdigão, the NEWA experimental site in Portugal. We renamed it "PO", Portugal, to avoid confusion.

16. P9 Tab2: It would help the reader if the acronyms (particularly the meaning of S1 and W1) was explained in the text/caption.

    Agreed. A short explanation has been added.

17. P8 L26: "the largest differences arise from the choice of PBL scheme, as shown in Fig 4". While the figure clearly illustrates the point that the authors make about the coincidence of regions with high surface roughness with areas of large differences between PBL schemes, it does not actually show, as the authors claim, that this is the largest difference. But even if it's not the largest difference, it would still be interesting to also show/quantify the effect of the different initialization strategy. Moreover, in the light of the authors' excellent point about the necessity to quantify differences between distributions, I'm quite surprised that they opted here to show the difference in the mean annual wind speed, rather than the more comprehensive EMD.

    Agreed. The figure does not show that the largest differences arise from the choice of PBL scheme. But, the following figures do. Except for the northwest of the NW domain and mountainous areas in the Pyrenees and the western Alps, the differences are larger in the MYNN-YSU than the W1-S1 comparison. We have toned down the statement in the manuscript.

    Instead of including more maps, which were not used in the original work, Figure 6 in the manuscript now shows the various statistics for the sites, including EMD. At the time this analysis was done we had not yet discovered the advantages of using the EMD.

18. P11 L4: I'm a bit concerned about the authors' conclusion that the weakly nudged setup is actually the best choice. Particularly, I would like to see whether the evaluation statistics are dependent on the lead time of the simulation.

    The plots depicting error metrics as a function of lead time are depicted in Figure 2. The lead times were aggregated into 12-hour bins, and the weighted average over all the stations is shown, with weights being the number of samples available for each bin in each station. The metrics shown are BIAS, the absolute value of BIAS, with the absolute value being taken before the averaging process, RMSE and EMD. No specific pattern can be observed that would describe how the error

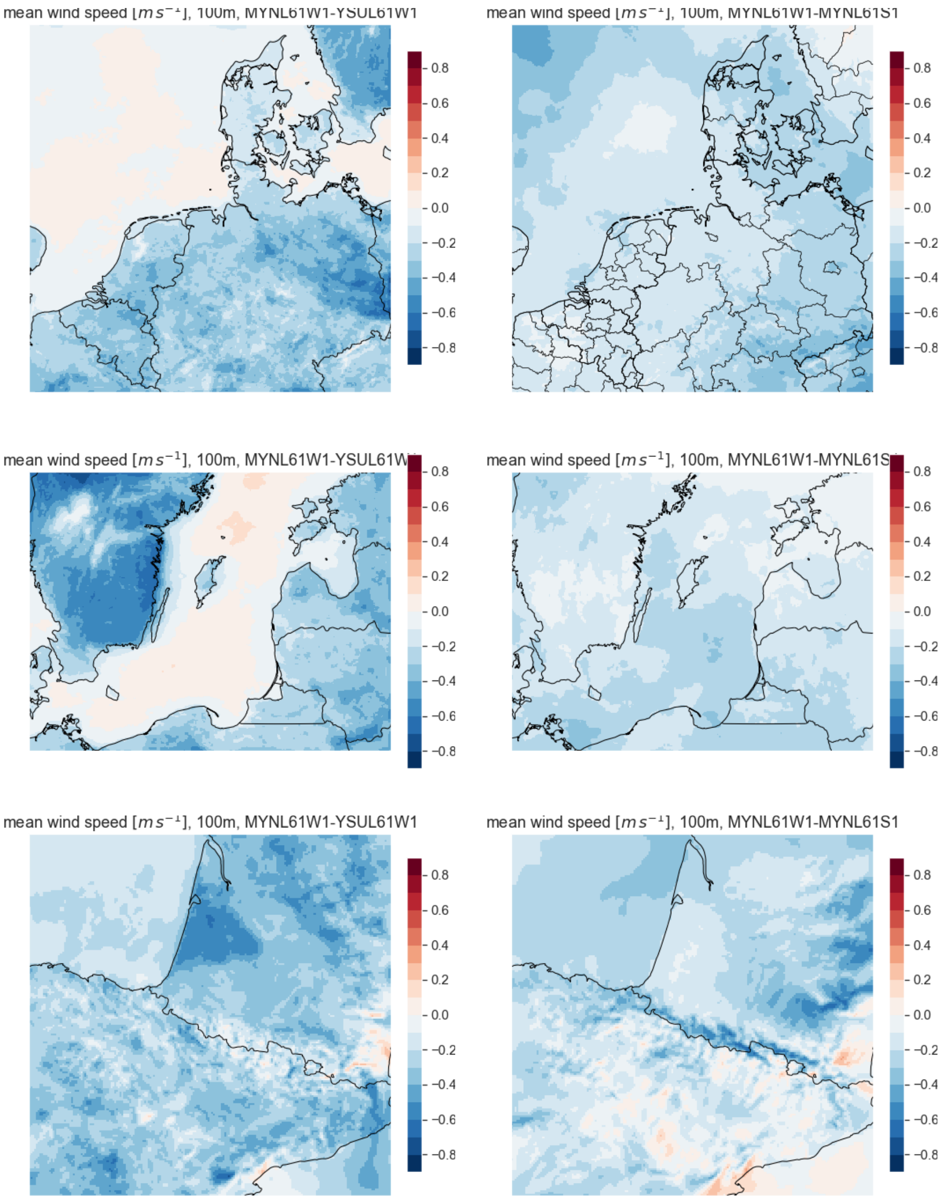


Figure 1: Differences in annual mean wind speed at 100 m between pairs of model simulations: MYNN-YSU on the left, W1-S1 on the right for three of the model domains. The colour bar is identical for all panels.

metrics evolve over time — probably the number of samples is too small and random errors dominate the distribution. On average, the error metrics for weekly runs are smaller than for the daily runs, which confirms the results described in the paper. The EMD for weekly runs is about the same as for the daily runs for MYNN PBL scheme, but YSU weekly runs seem to be associated with slight increase in EMD. However, one must take into account the difference in number of samples in each distribution, and it is likely, that EMD penalises the inhomogeneity that arises from the smaller number of samples. The figure shows a slight downward trend for both the BIAS and absolute value of BIAS, consistent with the hypothesis that the increased performance of weekly runs comes from the fact that the model solution is allowed to fully develop the mesoscale circulations, however, more detailed investigation of this matter is beyond the scope of this paper. The effect of the spin-up time was previously studied in Hahmann et al (2015) [5] and Vincent and Hahmann (2015) [9].

In conclusion, based on the statistics presented in this paper and previous studies, we believe the choice of the weekly simulations is justified. From the answer to Referee #1 (item number 12 in P8,L16) section 5.1 is now split into two sections. In the second of these, we have added a couple of sentences justifying the use of the weekly setup.

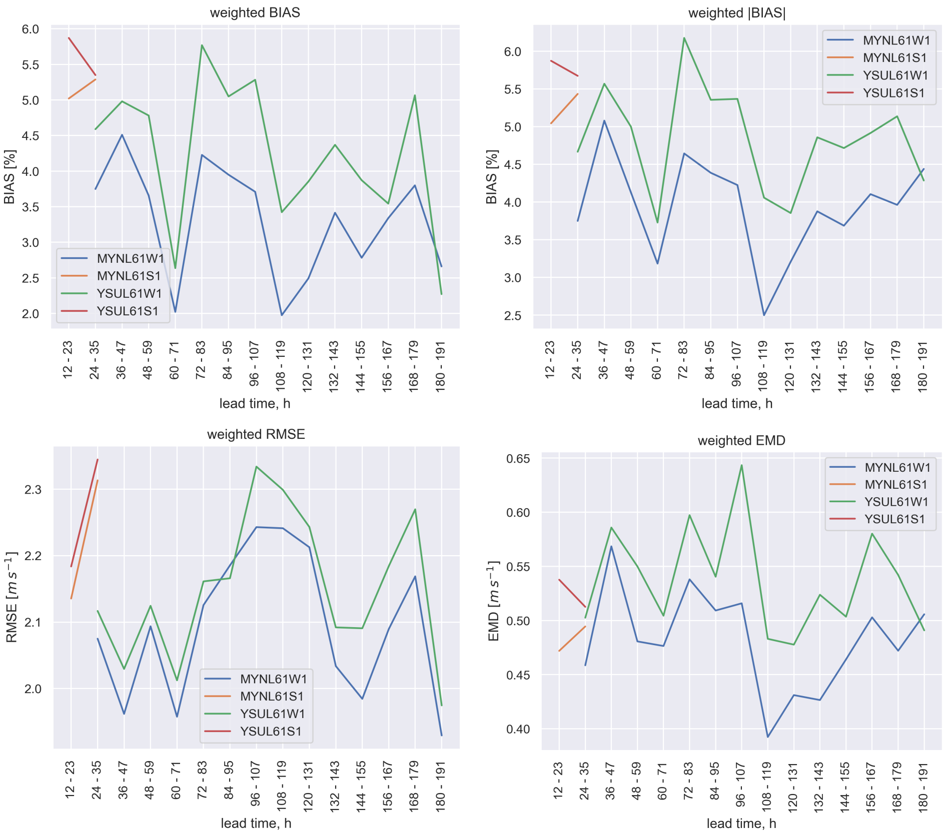


Figure 2: Error metrics as a function of lead time. Lead time is aggregated in 12-hour bins. Weighted average over all stations used in the analysis is shown with weights being the number of samples. Please note that the number of samples in each lead time bin for YSUL61S1 and MYNL61S1 runs is much larger than for YSUL61W1 and MYNL61W1, i.e. ∼ 4000 samples for S1 run bins and ∼ 600 samples for W1 bins.

19. P11 L9: Change title? Most of the section is about the modifications to MYNN.

    The title of the section heading has now been changed to “Sensitivity to properties of the MYNN scheme”

20. P12 Fig6: Is this figure for all mast heights? And are the differences shown here actually significant? Especially the correlation seems very consistent between all runs. And what about the earth mover’s distance? Why is it not shown here? Is the bar plot really the best choice here, seeing that differences are amplified or dampened depending on the choice of the axes’ intersection?

    No, the figure shows the metrics for the underlined heights in Table 1. We chose to validate for the height that is closest to the common turbine height $\sim 100\,\mathrm{m}$. This info has now been added to the figure caption.

    This figure has been replaced by a new figure that now includes the EMD and CEMD in addition to the BIAS and RMSE. This homogenises the results with the rest of the paper. However, at the time this analysis was done we had not yet discovered the advantages of using the EMD and the choice of model configuration was not based on this measure.

21. P13 L5: It would be useful to describe how these 25 configurations where selected from on the thousands of combinations alluded to before. Perhaps repeat or elaborate on the “expert judgement” here.

    For PBL/SL/LSM parameterisation the number of options is more finite because some combinations are technically not possible. At the beginning of our study, the table of experiments contained many more combinations. As mentioned in the following paragraph, many of the combinations simply did not run despite our attempts to do so and were excluded form the final set of experiments.

22. P16 L16: “absolute difference in relative bias”. This formulation is incorrect. A correct formulation would be “Fig 9b shows the difference in absolute relative bias between ...”.

    Agreed. Please see answer to item 7.

23. P17 Fig9: I would suggest to group figures 9 and 10 together, OR, to present 9a and 10a together, and 9b and 10b. As it is now, it is difficult to compare figures 9b and 10b. Also, consider using a different colormap for a and b, since right now green means ”good” in b, but not in a, in both figures.

    Thank you. This is a good suggestion. We have also homogenised all the figures that rely on “heatmaps”. Now any purple values are “good” and any brown values are “bad”. The new colour table is colour blind friendly.

24. P18 Fig10: It is not clear to me how the “relative” EMD is calculated in panel A. And is the same ”relative” EMD used for panel B? Why not just show the EMD in m/s? I feel the author”s are making things needlessly complicated. Same

question applies to the "absolute bias". Although I can see that the "non-relative" metrics are wind-speed dependent, mean wind speeds are all around 10 m/s, so the differences between sites will be very small. Therefore I would argue: simpler is better.

In total agreement. Please see response above (item 23).

25. P18 L5: I'm not sure if the choice for MO is justified based on the statistics shown. Although the EMD improves slightly for four sites, it degrades severely for some of the others. I'm not sure of the overall effect is positive. This could use some extra discussion.

    Agreed. This topic deserves a longer explanation

    We can start explaining what we mean by the best model setup. The best model setup would have the best verification metrics for all the stations analysed. The problem is that from the sensitivity analysis results it is clear that some model setups have better scores for one verification metric (e.g. EMD) and some setups have better scores in other metrics (e.g. RMSE). Also, the performance of the setup varies considerably from station to station. In addition, one can clearly see that performance at each station is systematic. There are "good" stations where all setups perform well (e.g. FINO1) and "bad" stations where all setups struggle (e.g. Risø). Should the model setup performance count more in the good stations? Or in the bad stations? The model bad performance is often associated with details of the station siting, which are poorly represented in the WRF model resolution used.

    In addition, our goal is to choose the model setup that would perform best over the full NEWA domain. Basing this choice on 8 stations introduces significant uncertainty. Taking into account all the results available to us (not all of them included in the manuscript due to the issues of space) and the limitations described above the conclusion is, that there is no single setup that could be easily identified as "the best". Instead, we have a small set of setups where each performs equally good (or bad) depending on the metric or station. An argument could be made that each of them should be used for the final product. While working on the project, we made the conscious decision not to delegate the decision to a simple algebra of taking the average metrics over all the stations, because that would introduce the assumption that this decision would not change, if for example another station is added.

    We would argue that we cannot distinguish the performance of these "good" members based on the observational data available. In addition, due to the limited computational resources available we had to look at the computation aspects of the setups. The MO setup is one of the best performing, according to the verification results. It performs well both in terms of distribution (EMD) and in time-series (RMSE). It has the additional benefit of runs being numerically stable, when compared to BASE, i.e., the runs failed less (this aspect was important, because of the necessity for a person to monitor and re-submit the runs), and MO also had a favourably small computational time when compared to other setups.

We are aware that arguments could be made that some other setup should have been chosen. We would have liked to be able to make this decision using more observations or after some additional analysis, however, due to the practical constrains, a decision had to be made based on what can only be described as “imperfect information”, a situation that might be familiar to many readers of this manuscript. Therefore, we would like to argue that the our choice is validated by the good evaluation results described in Part 2. We would like argue that although we are not claiming that we made the ”perfect” choice, that is, it is possible that choosing some other setup would have yielded even better metrics, but our choice was ”good enough” given the information we had available at the time.

In conclusion. We plan to include some of the points raised above in the revised manuscript especially in the “discussion and outlook” section.

26. P19 L8: This is interesting indeed. Perhaps the authors can discuss this observation a bit more in depth? I’m still not convinced that 8-day nudged simulations are the best choice.

    Agreed. We have expanded the figures to include the RMSE and CEMD. To follow this the discussion, we have expanded the discussion regarding the differences between statistics that look at the distribution versus time synchronisation.

    The new figures added show better agreement not only for the overall distribution (EMD), but also in the RMSE. The stronger nudging towards the reanalysis, which has observational data assimilated in it, in experiments NUDD3 and GNUD3 results in smaller RMSE. The interesting question is: why does it degrade the BIAS and EMD performance? A simple answer is that the WRF model mostly has increased performance over ERA5 (see Part 2 [1]) and therefore nudging towards the poorer performing model decreases model performance. A comprehensive answer to this question is beyond the scope of this paper, however, the conclusion is less surprising than it seems, taking into account the many different interactions between wind speed, surface energy budget, transport in the surface/planetary boundary layer, etc. Nudging is “artificial” or non-physical and can interfere with these complicated processes in non-trivial way.

27. P20 Fig11: Same comments as for fig 9 and 10: it would be better to use a different colormap for figures a and b, and perhaps group all figures together to prevent them spreading over multiple pages. Also reconsider using relative/normalized metrics.

    Indeed. The colour tables used in the figures have been homogenised and we now use the SS statistic to compare to the base simulation. Please see our answer to 23.

28. P20 L6: “at hub height”. Does this mean that only ∼100 m was used for all tables? So far I wasn’t sure, but I was under the impression that the metrics were calculated on the basis of all measurement heights. What does this mean for the representation of the (distribution of) wind shear between the various model simulations? I know that the mean profiles have very similar shear, but beware that instantaneous profiles can show substantial variation!

    All the statistics have been computed for the levels underlined in Table 1. As the reviewer suggests, the behaviour of the wind statistics could be very different at e.g. 10 meters. We focus on heights relevant for wind energy development. We have added the height of the validation to all figure captions.

    For lack of space we concentrate on a single level and do not consider the shear distributions. These have been analysed in a previous paper [5].

29. P20 L7: “Unfortunately, we did not run... so we cannot”. This statement contributes substantially to the overall impression that this manuscript is an accomplished fact.

    The sentence has been rewritten. However, it is a fact that the project is complete, and while some further simulations would have been very interesting to do, they were not able to be accomplished during the period of this study.

30. P21 L3: It seems a bit weird that this is the last experiment. If I would have designed this experiment, it would have been the first, as the other settings may depend on it. Especially the combination of domain size and nudging/initialization strategy seems influential.

    Yes, agreed. This was not the last experiment to be done. It was done at a similar time than the NW domain sensitivities. However, it seemed to fit better here for the flow of the manuscript. In addition, the production run did not follow the advice of using smaller domains. We just hope that the results described here are useful to someone else in the future.

31. P21 L8: This is an interesting dilemma. Did the authors modify the WRF registry to output only the relevant parameters? Would the ”restart” option not lift this constraint as the simulation time could be shortened to enable intermediate postprocessing? And how does the pan-European domain compare to the CONUS domain used in the rapid refresh configuration of NOAA? Have the authors contacted them for advise about their reference setup and HPC strategy? Options to stream the WRF output, or to access model fields during a simulation to postprocess them right away would be very welcome recommendations for model development. I wonder whether such features are already available, for example through the “basic model interface” developed by CSDMS.

    The creation of the simulations and the postprocessing of the output for a wind atlas is quite different from that of NWP output. For example, the histograms of combined wind speed and wind direction distribution are need for each model grid point and height for the complete duration of the model simulation. There are more details on the post-processing in the companion paper [1].

32. P21 L11: “outside region of interest . . . would be wasted”. I have to disagree here. Although it is not the explicit goal of the NEWA project, these data could be very useful for those non-EU countries. Again, please broaden the scope from ”NEWA” to ”a relevant and interesting dataset for the audience of GMD”. I think this dataset can have more impact if it would be available for other researchers as well. The term ”waste” therefore rubs me the wrong way.

    Agreed. The sentence was a bit harsh and has been rewritten. For context, the computational and time resources were limited — the simulations took 6 months to compute, and every available CPU hour allocated in the PRACE grant for using the cluster at the Barcelona Supercomputer centre (BSC) was used. Also, the data storage is nearly 160 TB, which took almost as long to transfer from the BSC as it did to compute. The scarcity of resources meant that the resources available had to be used frugally. Therefore, while we would have liked to extend the spatial coverage and scientific questions we could answer, hard decisions had to be made to prioritise what could be feasibly handled given the resources available.

33. P21 L26: (and possibly . . . not shown). Model runs that would show this have also not been described as far as I can see. What additional simulations did the authors perform that inspire this statement, or is it mere speculation?

    A simulation with a large outer domain and a small inner domain was carried out. The mean wind speed from this extra simulation resembles that of the LG simulation more than the SM simulation. From this, we infer that the position of the inflow boundary is important, but more research is needed. To keep the discussion brief this simulation was not added to the original manuscript. We have rewritten the sentence in the updated manuscript removing the mention of additional simulations that have not been explained.

34. P21 L34: “We decided, however, against very small domains. In terms of accuracy they would probably perform better”. Not only does this statement sound speculative, it also partly undermines the objective of the paper. If one of the options considered (the SM domains) was not an option to begin with, why test it? For some sites, the impact of this decision seems to be larger than the accuracy gained through the detailed optimization of all other settings of the model. . .

    Yes, agreed. It is contradictory, but we hope that this would be useful information to future wind modellers. Please see our answer to question 30.

35. P22 Fig13: The y-axis is unreadable.

Apologies. The figure has now been replaced with the heatmaps tables to match previous discussion.

36. P24 L2-3: "In that paper we conclude that..." ? Better than just using "raw" ERA5 data?

    Agreed. The sentence is too strong and perhaps not relevant here. It has been rewritten.

37. P24 L4: "some questions remain unresolved ... expensive nature of the numerical experiments". This is obviously true, but I feel there are many more questions unanswered because of limited manpower. I'd really appreciate it if the authors could reflect more on that aspect of their study.

    Agreed. Many questions always remain unanswered. And because of the limited set of observations this was not the best region to carry out the many test needed. However, this study still has done much more than previous wind atlas studies. The project has ended and there is limited funding to continue the analysis of the results.

    Some issues regarding the diurnal cycle in the observations and the model simulations are currently being studied. We hope a new publication will result from that analysis.

38. P24 L17: "It would have been optimal..." again this contributes to the "accomplished fact" feeling. This can still be done, can't it? And it can answer some of the questions I have asked, e.g. P5 Fig1 related to the representativeness of the northern domains for Southern Europe.

    Unfortunately this is not possible, due to the limited availability of relevant datasets for validation. However, Part 2 [1] of the study shows very good comparison of the WRF simulations using the final configuration against the Vestas tall tower data. However, because the data is proprietary, the geographical distribution of the errors cannot be shown. Figure 10 of that study shows larger errors in, for example, Turkey compared to the northern sites, but the larger number of sites in France does not show the same tendencies. In addition, we have added to the revised manuscript that "This large dataset can be further verified as additional data becomes available". Hopefully this helps alleviate the "accomplished fact".

    In the revised manuscript we have strengthen the connection between the two parts of the study. This should help clarify many aspects of part 1.

39. P24: The discussion (or other parts of the paper if appropriate) should also address why vertical resolution was not subject to sensitivity analysis, what the uncertainty of the observations is, why wind direction is not considered at all, whether performance is similar across different heights, why/how wind shear has

(not) been assessed, how the set-up compares to other similar efforts. The outlook should offer some advice for future studies: what have we learned from this study, in what direction should model development evolve, what are the main strengths/weaknesses of the WRF setup, which parameterisation schemes should we abandon right away, etc.

Agreed. Further discussion has been added. In early experiments, not reported in the manuscript, we experimented with increasing the number of vertical levels from 61 to 91. The results of that experiment showed very small differences between the simulations.

# References

[1] Martin Dörenkämper et al. "The Production of the New European Wind Atlas, Part 2: Production and Validation". In: *Geosci. Model Dev. Discuss.* in review (2020).

[2] Caroline Draxl et al. "Evaluating winds and vertical wind shear from WRF model forecasts using seven PBL schemes". In: *Wind Energy* 17 (2014), pp. 39–55. DOI: `10.1002/we.1555`.

[3] Filippo Giorgi. "Thirty Years of Regional Climate Modeling: Where Are We and Where Are We Going next?" In: *J. Geophys. Res. Atmos.* (June 2019), 2018JD030094. ISSN: 2169-897X. DOI: `10.1029/2018JD030094`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1029/2018JD030094`.

[4] Andrea N. Hahmann et al. *Mesoscale modeling for the wind atlas for South Africa (WASA) Project*. Tech. rep. DTU Wind Energy, p. 77. URL: `http://orbit.dtu.dk/services/downloadRegister/107110172/DTU_Wind_Energy_E_0050.pdf`.

[5] Andrea N Hahmann et al. "Wind climate estimation using WRF model output: Method and model sensitivities over the sea". In: *international Journal of Climatology* 35 (2015), pp. 3422–3439. DOI: `10.1002/joc.4217`.

[6] Noam Lupu, Lucía Selios, Zach Warner, et al. "A new measure of congruence: The Earth Mover's Distance". In: *Political Analysis* 25.1 (2017), pp. 95–113.

[7] Julien Rabin, Julie Delon, and Yann Gousseau. "Circular Earth Mover's Distance for the comparison of local features". In: *2008 19th Int. Conf. Pattern Recognit.* IEEE, Dec. 2008, pp. 1–4. ISBN: 978-1-4244-2174-9. DOI: `10.1109/ICPR.2008.4761372`. URL: `http://10.0.4.85/icpr.2008.4761372%20https://dx.doi.org/10.1109/ICPR.2008.4761372%20http://ieeexplore.ieee.org/document/4761372/`.

[8] J. Ramon et al. "The Tall Tower Dataset: a unique initiative to boost wind energy research". In: *Earth System Science Data* 12.1 (2020), pp. 429–439. DOI: `10.5194/essd-12-429-2020`. URL: `https://essd.copernicus.org/articles/12/429/2020/`.

[9] Claire L. Vincent and A. N. Hahmann. “The Impact of Grid and Spectral Nudging on the Variance of the Near-Surface Wind Speed”. In: *Journal of Applied Meteorology and Climatology* 54 (2015), pp. 1021–1038. DOI: `10.1175/JAMC-D-14-0047.1`.

## Editorial remarks

1. Excessive use of commas and conjunctions make parts of the text difficult to read. this can easily be addressed by making shorter sentences. For example:

   Yes, we agree. We have revised the manuscript and shortened sentences whenever possible. The specific sentences have been corrected as listed below.

   - P2 L6-7: rewrite ”but not only”. Very wordy sentence, it has been rewritten.
   - P2 L7-8: use only on of ”for example ... to name a few”. Fixed.
   - P2 L13-15: suggest ” ... its usefulness. It has ...”. Done.
   - P3 L13-16: start new sentence at ”however”. Done.
   - P3 L10: start new sentence at ”however”. Done.
   - P3 L12: ”A large number” or ”Large numbers of” . Done. We have also removed unnecessary words from the sentence.
   - P3 L16-18: move ”has been reported” to beginning: ”a number of studies report ...”. Done.
   - P3 L19: remove comma after ”cases”, suggest: ”two processes with opposing effects” (remove ”canceling each other out”) Done.
   - P3 L21: citation without brackets. Fixed
   - P3 L28: coastal winds? Flow is ambiguous (air or water). We clarified that we refer to the atmospheric flow in the coastal zone.
   - P4 L8: Simulations (plural), or perhaps ”reference configuration”? Yes, it should be simulations.
   - P6 L23: remove ”in”. Removed.
   - P7 L12: suggest: ”Small changes in wind speed are (thus) amplified when converted to power.” Done.
   - P8 L18-20: suggest to split in 2 or 3 shorter sentences. Remove ”the aim was”, as the next sentence also states ”the objective”. Done.
   - P8 L20: remove ”left” (or write ”left untouched”?) Done. Agreed. This was a very complex sentence. It was been simplified by using parenthesis.
   - P8 L22: ”or if there were regional differences” can be omitted as it is already implied by the use of ”whether”. Done.
   - P10 L4: better to split up and rephrase, instead of using ”but” twice in the same sentence. Done.

- P10 L6: this sentence can also be split in two shorter sentences.
- P11 L6: Unclear, long sentence. Done.
- P15 L2: "regional" should be "region(s)"? Done.
- P17 L20: "conclusions can be drawn". Done.
- P18 L1: "scheme and run" both refer to a scheme/set-up/configuration, right? Yes, the name "simulations" has been used instead.
- P21 L21: "six" instead of 6 (in line with the surrounding text) Done.
- P21 L34: rephrase "which would face". The two sentences have been rewritten
- P22 L8: unclear sentence; a.g.l. and AGL are the same. Which figures? Indeed. Very confusing sentence. It has been rewritten.
- P23 L18: weird use of commas around ". . . change source...". Done. Should have been "changing the source..."
- P24 L2: "wind climate". Done.
- P24 L6-7: "however... but ...". Done.
- P24 L10: "best optimal" . Missing verb. Fixed.
- P24 L17: "observational dataset". Added.

# Response to Referee #2

Thank you for the comprehensive comments. Your comments, together with those of referee #1, led to a thorough revision of the paper.

The most general comments regarding the revisions to the manuscript are:

1. At the start of the research project typically there are high expectations placed on the sensitivity experiments, however, reality always brings some corrections and caveats. Given the enormous possibilities in setting up WRF, an "optimal" configuration is unreachable. We have tried to revise the introduction to convey that the paper focuses on finding the "best possible" model configuration **constrained** by the practical issues in running the model simulations and the ultimate goal to use the simulations for a **wind atlas**.

2. We humbly think that the present analysis offers and unprecedented assessment in terms of the variety of WRF model configurations tested.

3. The manuscript aims to tell the story of how the NEWA wind atlas came to be. Therefore, further analysis of the model results will make the flow of the paper less clear. We have tried to enhance this structure in the revised manuscript.

4. We have replaced some of the figures (6, 9–12, 13) to homogenise the analysis of the results. We have also added new figures including the RMSE and circular EMD for wind direction.

5. We strengthened the connection to the companion paper, https://www.geosci-model-dev-discuss.net/gmd-2020-23/, which is now available.

The reviewers' comments are in black and our responses in blue.

# General Comments

1. 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 complimentary metric alongside the typical timeseries-based performance metrics.

   Thank you. As described above, we have expanded the use of the EMD and Circular EMD (CEMD) for wind direction in the manuscript.

2. 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 collaboration 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.

In the second part of this study [1], the final wind atlas is validated against masts over all of Europe. However, at the time we did the sensitivity simulations and we needed to decide on a final configuration, further evaluation with data besides the 8 sites in N. Europe was not possible. The public data from tall masts needed for evaluation are scarce over Europe. In a recent paper and database [8], where a global database of tall masts was compiled, there is only a handful of mast over Europe where data are available and only a couple lie in the region chosen for the sensitivity study. Further, even if the data used in part 2 would have been available, the evaluation would not have been possible. There are no masts in Denmark and very few in Germany and the masts in Poland have data for only a few months in 2015. (see Figure 4 of Part 2)

3. 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.

   Agreed. The new manuscript includes further figures with all the statistics including BIAS, RMSE, EMD for wind speed and CEMD for the wind direction.

4. 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?

   There seems to be some confusion here and we have clarified the forcing data used in the manuscript. The sensitivity simulations described in sections 5.1, 5.2 and 5.5 used ERA-Interim as forcing. All other simulations (excluding the simulation named "ERAI" in section 5.4) used ERA5 data. Actually the sensitivity test of replacing ERA5 with ERA-Interim is described in Table 5 and the results are shown in the heatmap plots. The differences in BIAS and EMD (figures 11 and 12 in the manuscript) show very small differences. Since ERA-Interim was scheduled to be discontinued in 2019, we chose to continue our sensitivity studies and production run with the ERA5 dataset.

5. 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.

   Thank you. We believe the document is much improved after this round of revisions.

## Specific Comments

1. 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 wind climates, should be discussed.

   The sensitivity analysis was done only for the domain over Northern Europe. As mentioned in our answer to item 2 above, tall masts of good quality data publicly available are very scarce. The implications are the focus of section 5.1, where we argue that the behaviour of the mean wind speed relative to various PBL/SL parameterisations is similar among the five domains in very distinct wind climates.

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

   The paragraph where this statement appears has now been rewritten. We refer here to the whole wind atlas method, which is now described in a little more detail. Therefore, "linearised model" is now "linearised method".

3. 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.

   This is now better explained in section 5.1. It is a limitation of the study, but we had no alternatives. In retrospect, the validation in paper 2 [1] shows that the resulting wind atlas provides good estimates not only in N. Europe, but also in other regions.

4. 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.

We have added a description of the temporal resolution of the data, which was 10-min means that were filtered to hourly using the period closest to the top of the hour. Additionally, Table 1 has been extended to include wind directions, and show the data availability as a percentage rather than number of samples.

5. 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.

   We now explain in the text that the impact of the wind farm is difficult to quantify. For example, at some of the sites, wind farms were being built and tested in 2015, and without operational data, we cannot know when a wind farm was curtailed or otherwise not operating. Additionally, filtering for the wind farm possible perturbation can severely decrease the number of samples for some sites. We have now added the centre of the filtering wind direction and the fact that there is an additional wind farm near the FINO2 mast. The text in the revised manuscript has been changed to "...the data has not been filtered or corrected for the turbine wakes. However, the presence of the wind farm can impact the evaluation of the model results and should be kept in mind."

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

   Agreed. We have replaced "understood" by "interpreted"

7. 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.

   Agreed. This is a very good suggestion. A new panel has been added to Figure 2 showing the cumulative distribution functions.

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

   Circular variables are variables, like wind direction, where there is an apparent discontinuity at between 0 and 360°. In the updated manuscript we use a version of EMD metric adapted for circular variables (CEMD), which was used to evaluate the wind direction distributions in the model simulations.

9. 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?

   At the start of the project (summer 2015), WRF V3.6.1 was not that old (it was released August 14, 2014) and it provided good evaluation against observations in other regions, for example South Africa [4]. Using the latest version of a model is not always advantageous as seen in the changes to the MYNN parameterisation in WRF V3.8.1, that heavily impacted the validation statistics. Later in the large ensemble we moved to WRF V3.8.1. We acknowledge that the model version used in the various simulations was not clearly stated. This situation is fixed in the revised manuscript.

10. 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?

    What was meant by the statement was that normally the winds in the YSU scheme are larger than those in the MYNN scheme (see Figure 1 in the answer to the comments from reviewer #2). But when conditions are mostly unstable, as it is in the French Atlantic coast (50–60% of the time), Mediterranean sea (60–70% of the time) and some coastal areas Turkey, the situation reverses and the 100m mean winds are higher in the simulations using the MYNN scheme than the YSU scheme. Yes, conditions are mostly stable or neutral over the North Sea and the Baltic Sea. The sentence in the text has been expanded to clarify this issue.

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

    Totally agree. At the time that these analyses were originally made we had yet to discover the advantages of the EMD metric. But now we show this metric throughout the manuscript and also in Figure 6.

12. 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.

    The objective of the figure was to show that the wind profiles from the simulations clustered, not to be able to differentiate between them. We have redone the figure with a single grey colour, highlighting only the results from two relevant simulations. Hopefully it will reflect better our intention.

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

    Thanks. It is a good suggestion. However, in this case we would argue against adding the averaging over the stations. Please see the long discussion in the response to Referee #1, item 25 (P18 L5). We don't want to give the impression that the decision on final configuration was just based on a raw evaluation of the numbers. Adding the average over the stations will convey that impression in our opinion.

14. 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.

    Thank you, this is an important question. As we searched for the "best" model configuration, we kept asking "Is there another different configuration that will be better than our base?" The relative heatmaps help answer that, while also showing how small the differences between simulations are, which is sometimes hard to spot in the BIAS or EMD plots alone. This is because the differences between the stations are often more pronounced, for absolute values of metrics, than the differences between the ensemble members at the same station. It is important to note, that this method of examining results does not assume that the "BASE" setup is the most accurate, it is just a more convenient way of identifying differences between different models.

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

    Please see our reasoning above.

16. 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 analysis needs to be provided.

    Agreed. It is a very fair question. In conclusion the simulations show that many parameters usually though to be important for NWP or climate modelling have little or no influence. So it would probably be fair to choose any of them, except for some PBL/LS/LSM combinations that definitely degrade the validation metrics. A long discussion on the matter was given in the answer to referee #1 (item 25, P18, L5). In the revised manuscript we try to convey this in a more direct manner.

17. 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?

All experiments in these tables used ERA5. The comparison is then from ERA5 to MERRA2, FNL, and ERAI. As the table reveals the differences between ERAI and ERA5 are small.

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

    All the experiments, except for ERAI, MERRA2 and FNL, were carried out using ERA5.

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

    Same response as above. Sorry about the confusion.

# The Making of the New European Wind Atlas, Part 1: Model Sensitivity

Andrea N. Hahmann[1], Tija Sīle[2], Björn Witha[3,4], Neil N. Davis[1], Martin Dörenkämper[5], Yasemin Ezber[6], Elena García-Bustamante[7], J. Fidel González-Rouco[8,9], Jorge Navarro[7], Bjarke T. Olsen[1], and Stefan Söderberg[10,11]

[1]Wind Energy Department, Technical University of Denmark, Roskilde, Denmark
[2]Institute of Numerical Modelling, Department of Physics, University of Latvia, Riga, Latvia
[3]ForWind, Carl von Ossietzky University Oldenburg, Germany
[4]energy & meteo systems GmbH, Oldenburg, Germany
[5]Fraunhofer Institute for Wind Energy Systems, Oldenburg, Germany
[6]Eurasia Institute of Earth Sciences, Istanbul Technical University, Istanbul, Turkey
[7]Wind Energy Unit, CIEMAT, Madrid, Spain
[8]Dept. of Earth Physics and Astrophysics, University Complutense of Madrid, Madrid, Spain
[9]Institute of Geosciences, IGEO (UCM-CSIC), Madrid, Spain
[10]WeatherTech, Sweden
[11]Renewable Energy Analytics, DNV-GL Energy, Sweden

**Correspondence:** Andrea N. Hahmann (ahah@dtu.dk)

**Abstract.** This is the first of two papers that ~~documents~~ document the creation of the New European Wind Atlas (NEWA). It describes the sensitivity analysis and evaluation procedures that formed the basis for choosing the final setup of the mesoscale model simulations of the wind atlas. ~~An optimal~~ The suitable combination of model setup and parameterisations, bound by practical constraints, was found for simulating the climatology of the wind field at turbine-relevant heights with the Weather Research and Forecasting (WRF) model. Initial WRF model sensitivity experiments compared the wind climate generated by using two commonly used planetary boundary layer schemes and were carried out over several regions in Europe. They confirmed that the ~~largest~~ most significant differences in annual mean wind speed at 100 m above ground level mostly coincide with areas of high surface roughness length and not with the location of the domains or maximum wind speed. Then an ensemble of more than 50 simulations with different setups for a single year ~~were~~ was carried out for one domain covering Northern Europe, for which tall mast observations were available. ~~Many different parameters were varied~~ We varied many different parameters across the simulations, for example, model version, forcing data, various physical parameterisations and the size of the model domain. These simulations showed that although virtually every parameter change affects the results in some way, significant changes on the wind climate in the boundary layer are mostly due to using different physical parameterisations, especially the planetary boundary layer scheme, the representation of the land surface, and the prescribed surface roughness length. Also, the setup of the simulations, such as the integration length and the domain size, can considerably influence the results. ~~The~~ We assessed the degree of similarity between winds simulated by the WRF ensemble members and the observations ~~was assessed~~ using a suite of metrics, including the Earth Mover's Distance (EMD), a statistic that measures the distance between two probability distributions. The EMD was used to diagnose the performance of each ensemble member using the

full wind speed and direction distribution, which is ~~important~~ essential for wind resource assessment. ~~The~~ We identified the most realistic ensemble members ~~were identified~~ to determine the most suitable configuration to be used in the final production run, which is fully described and evaluated in the second part of this study (Dörenkämper et al., 2020).

*Copyright statement.*

## 1 Introduction

Wind atlases ~~can be~~ are defined as databases of wind speed and direction statistics at several heights in the planetary boundary layer. ~~These~~ Wind atlases have been created for many regions of the world ~~mainly~~ to help inform wind energy installations, but ~~not only,~~ many other human activities benefit from the knowledge of the wind behaviour at its climatology. For example, ~~for~~ wind atlases can be used in structural design for buildings, transportation infrastructure and operation, recreation and tourism~~, to name a few~~. In 1989, the European Wind Atlas (EWA, Troen and Petersen, 1989) was released, which provided the first ~~wind atlas covering all of Europe~~ public source of wind climate data that covered the whole of Europe in a homogeneous way. The EWA was ~~mostly based on surface observations~~based on data from around 200 meteorological stations, and used the so-called ~~Wind Atlas method (Troen and Petersen, 1989), which makes it possible to transfer detailed information about the mean wind climate~~ *wind atlas method*, a collection of statistical models that are the core of the Wind Atlas Analysis and Application Program (WAsP) software package (Mortensen et al., 2011). This method allows for the transfer of the wind climate statistics from one location to another ~~. Before the New European Wind Atlas (NEWA[a], Petersen, 2017), the EWA was the only public source of wind climate data that covered the whole of Europe in a homogeneous way. Given that the~~ based on the surface characteristics. The EWA is now 30 years old ~~, it is lacking information that limits it usefulness, that is it~~ and of limited usefulness. It has a very coarse spatial resolution, does not provide time series of the variables of interest, and was developed using a linearised ~~model~~method (Troen and Petersen, 1989), which limited its applicability in complex terrain. ~~Nowadays, modern~~

In 2013, a team of 30 partners from 8 European countries started work on the “New European Wind Atlas” (NEWA) project (Petersen, 2017). The project had three main components: a series of intensive measuring campaigns, a thorough examination and redesign of the model chain (downscaling from global to mesoscale and to microscale models), and the creation of the wind atlas database. The scope of the modelling in NEWA is ambitious and too long for a single article. Therefore it is divided into two parts in this paper. The first one (this article) deals with sensitivity simulations and the second part (Dörenkämper et al., 2020) describes the production of the new database[b] and its evaluation.

Nowadays, wind atlases rely on the output from mesoscale model simulations, either sampling recurrent atmospheric states (e.g., Frank and Landberg, 1997; Pinard et al., 2005; Badger et al., 2014; Chávez-Arroyo et al., 2015) or by long-term simulations with Numerical Weather Prediction ~~(NWP)~~ models (e.g., Tammelin et al., 2013; Nawri et al., 2014; Hahmann et al.,

---

[b] https://map.neweuropeanwindatlas.eu/

2015; Draxl et al., 2015; Wijnant et al., 2019). NEWA follows the latter approach and provides a unified high-resolution and publicly available dataset of wind resource parameters covering all of Europe and Turkey. The wind atlas is based on 30 years of mesoscale model simulations with the Weather Research and Forecasting (WRF, Skamarock et al., 2008) model at 3 km × 3 km spatial and 30 min temporal resolution and 7 vertical levels. The WRF model was chosen because of its open access, it is used by many wind energy research institutes and private companies, and it was familiar to all the NEWA partners. Wind statistics from further downscaling with a microscale model (see Dörenkämper et al., 2020, for more details) are provided for Europe and Turkey onshore and offshore up to 100 km from the coastline, plus the Baltic and the North Seas with a horizontal grid spacing of 50 m at three wind-turbine relevant heights.

Mesoscale models are, in general, not specifically developed for wind energy applications; however, over the last decade they have been extensively used for that purpose (see Olsen et al., 2017, for a review). Developing an optimal WRF model configuration for wind resource assessment is not a straightforward task, considering the large number of degrees of freedom in the model configuration, and the different choices of input data. Among the configuration options offered in the WRF model are, physical parameterisations such as planetary boundary layer (PBL), surface layer (SL), land surface model (LSM), cloud micro-physics, and radiation. Also numerical and technical options (e.g., domain layout, nudging options, time step), and the initial and boundary conditions of the atmosphere, sea surface, and land surface are relevant aspects to be explored before determining the set up that better fits a specific application. ~~It is impossible to test every combination of these parameters, as the number of such experiments would be in the thousands, which is unfeasible in terms of computational resources. Therefore, a compromise between available computational power and scientific soundness had to be found. The approach in NEWA was to first define a “best practice” setup using the vast and diverse experience of the mesoscale modellers in the project, and then to test the sensitivity of the results to changes in the model configuration that are presumably the most relevant for the simulation of the wind field. This includes some physical options, such as PBL schemes, but also included a wide range of other parameters, such as numerical options, for which sensitivity results are rarely reported in the literature. Those parameters that did not evidence an impact in the simulation of the wind field where fixed as in the best practice set up. It is impossible to claim that all existing sensitivities in the model were found and tested, however the~~ Arguably, an optimal configuration that performs best at all time and spatial scales cannot be expected, and we search herein for a configuration that tends to perform better at most instances within the ensemble of sensitivity experiments performed.

A large number of ~~parameters that were tested and found not to be influential gives some credibility to our approach.~~

~~Large number of~~ ensembles of model simulations using the WRF model are documented in the literature~~in many applications~~. The WRF model is also used for more general climate research purposes as a regional climate model (RCM), for example, within the context of the Coordinated Regional Climate Downscaling Experiment (CORDEX) project (Katragkou et al., 2015)~~, however, ~~. However, in such cases the attention is typically focused on climate-relevant parameters, such as temperature or precipitation. A number of ~~sensitivity studies for~~ studies have reported sensitivity to the dependence of model-simulated temperature and precipitation on cloud microphysics, convection and radiation schemes (Katragkou et al., 2015), or PBL schemes (García-Díez et al., 2013) and all of the above (Strobach and Bel, 2019)~~has been reported. These show that the biases in model results can depend on the model setup, study region, season or diurnal cycle. Additionally, in some cases, it is suspected that~~

~~reduction of bias in a specific setup can be caused by errors in two different processes having the opposite effect and cancelling each other out. There are cases where all model setups fail to replicate some aspect correctly when compared to observations, e.g. as shown in (Mooney et al., 2017) for the diurnal cycle of precipitation.~~ . Sensitivity studies with a large number of WRF model simulations for wind energy applications have also been reported in Lee et al. (2012); Siuta et al. (2017); Fernández-González et al. (2017); Fernández-González et al. (2018) for ~~PBL wind~~ PBL-scheme ensemble prediction. Very few studies afforded an exhaustive sensitivity analysis of the model performance on the near-surface long-term wind climatology and many lack the verification at wind turbine heights. Two examples that looked at the sensitivity of the modelled wind climate are Hahmann et al. (2015), ~~who~~ that investigated a limited number of model parameters over the sea in Northern Europe and Floors et al. (2018b), ~~who~~ that concentrated on the impact of the model's spatial resolution on the ~~coastal flow~~ atmospheric flow in the coastal region. This study expands on these earlier attempts with a much larger set of sensitivity simulations and the comparison to ~~observations for the wide European domain.~~ a limited set of observations. It is impossible to test every combination of the WRF model setup and possible parameterisations, as the number of such experiments would be in the thousands, which is unfeasible in terms of computational resources. Therefore, a compromise between available computational power and scientific soundness had to be found. The approach in NEWA was to first define a "best practice" setup using the vast and diverse experience of the mesoscale modellers in the project, and then to test the sensitivity of the results to changes in the model configuration that are presumably the most relevant for the simulation of the wind field. This includes some physical options, such as PBL schemes, but also included a wide range of other parameters, such as numerical options, for which sensitivity results are rarely reported in the literature. Those parameters that did not show an impact in the simulation of the wind field were fixed as in the best practice set up. It is impossible to claim that all existing sensitivities in the model were found and tested. However the large number of parameters that were tested and found not to be influential gives some credibility to our approach.

~~With this background, our objective is to summarise the mesoscale simulations that form the backbone of the New European Wind Atlas. The scope of mesoscale modelling in NEWA is ambitious and cannot be addressed within one article. Therefore it is divided into two parts. The first one (this article) deals with sensitivity simulations and the second part (Dörenkämper et al., 2020) describes the production run and its evaluation.~~

All simulations in this study covered one full year and used similar grid parameters and modelling setup, which will be briefly described in Section 2. The data used for the evaluation of the ensemble of simulations among the whole pool of cases that are best suited to provide a meaningful sensitivity range is presented in Section 3; the statistics used in the model assessment and comparison among ensemble members is introduced in Section 4. The process of finding the most adequate (in a sense that will be defined) combination of model setup and parameterisations occurred in several phases: (1) analysis of sensitivity dependence on the geographical domain (Section 5.1), (2) selection of the WRF model version (Section 5.3), (3) creation of a large multi-physics ensemble (Sections 5.4 and 5.5) and (4) the analysis of the model sensitivity to the size of the model domain (Section 5.6). A summary of the findings of the sensitivity experiments can be found in Section 5.7. The paper ends with a discussion of the limitations of the approach used and the outlook (Section 6).

## 2 Description of the WRF model ~~simulation~~simulations

The database of simulated winds and wind-energy relevant parameters for the model sensitivity tests was created by splitting the simulation period into a series of relatively short WRF model runs that, after concatenation, cover at least a year. The simulations overlap in time during the spin-up period, typically between 3 to 24 h, which is discarded, as described in ~~(Hahmann et al., 2010, 2015; Jiménez et al., 2010, 2013)~~Hahmann et al. (2010, 2015); Jiménez et al. (2010, 2013). Two approaches are tested: frequent re-initialisation, in our case daily 36-hour runs, versus several days long runs that are nudged towards the forcing reanalysis. In the first approach, the re-initialisation every day keeps the runs close to the driving reanalysis and the model solution is free to develop its own internal variability. In the second approach, the use of nudging prevents the model solution from drifting from the observed large-scale atmospheric patterns, and the multi-days simulation ensures that the mesoscale flow is fully in equilibrium with the mesoscale characteristics of the terrain (Vincent and Hahmann, 2015). Both methods have the added advantage that the simulations are independent of each other, and therefore, can be computed in parallel, reducing the total time needed to complete a multi-year climatology. ~~In comparison, in continuous regional climate simulations~~ Alternatively, a single run ~~can last several wall clock months. Although these~~ of a continuous model experiment can be performed. Although this type of continuous ~~runs might~~ may present certain advantages, ~~as for instance that they preserve~~ like preserving the memory of land-atmosphere processes ~~, the selection of one approach or the other depends upon the needs of the specific experiment~~(e.g. Jiménez et al., 2011), its is not warranted more accurate and it can significantly expand wall clock time and make long-term high-resolution simulations not viable.

All mesoscale simulations in NEWA used three nested domains with a 3 km horizontal grid spacing for the innermost grid and a 1:3 ratio between inner and outer domain resolution, leading to 3 different resolutions: 27 km for the outer domain, and 9 km and 3 km for the inner nested domains. The model top was set to 50 ~~HPa~~hPa, following the best practices recommended by the WRF developers (Wang et al., 2019). The temporal coverage of the sensitivity simulations is one year (2015 ~~)~~or 2016), based on data availability. Other parameters common to all simulations are listed in Table A1. We explore the effect of changing various relevant parameters of the simulation set up from the base model configuration explained above to estimate the wind climatology over Europe.

## 3 Observed data

High-quality data from tall masts for the evaluation of the various sensitivity experiments is rare. In this study we used data from eight sites in northern Europe. The locations and names of the sites are shown in ~~Figure~~Fig 1; the details of the sites are summarised in Table 1. All sites are equipped with towers, and IJmuiden has an additional Zephir 300 continuous-wave lidar recording wind speed and direction at 25 m intervals between 90 and 315 m (Kalverla et al., 2017).

High quality data is essential for model evaluation (Lucio-Eceiza et al., 2018a, b). The mast data has been quality controlled and the mast flow distortion on the wind speed estimates was minimised by sub-setting the data. At FINO1, FINO2, Risø and Høvsøre, where wind speed measurements are available from only one boom, winds originating from ± 10° of the boom direction are filtered. A misalignment correction to the wind direction was applied to the wind vanes at FINO1

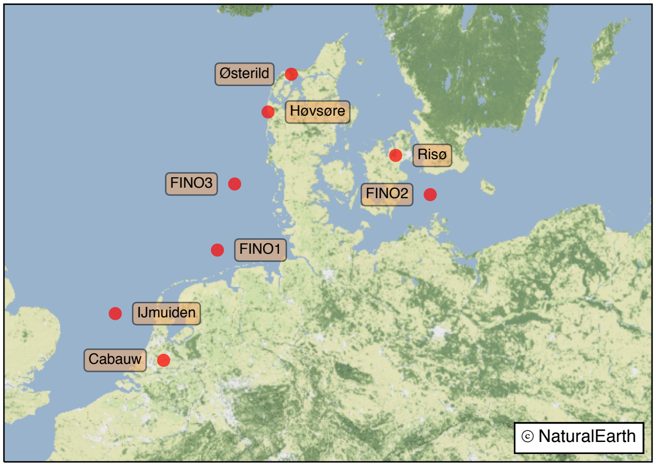


**Figure 1.** Location and name of the tall mast/lidar sites used in the model evaluation. Base map created with Natural Earth.

(Westerhellweg et al., 2012). At FINO3, we use the data from the three heights where wind speed measurements from three booms are available. The wind speed value is taken from the boom where the wind direction is most perpendicular to the boom direction. At IJmuiden, the data was processed as discussed in Kalverla et al. (2017). At Cabauw, the data was processed and gap filled as described in ~~(Bosveld, 2019)~~Bosveld (2019).

In addition to mast flow distortion, the wind speed estimates at some of the measurement sites are impacted by nearby wind farms. At Høvsøre and Østerild, test turbines are located north of the mast, in the sector with the least frequent wind directions. At FINO1, the wind farm *Alpha Ventus* impacts the eastern sector with the nearest turbine only 405 m away in the direction of 90°. The wind farm *EnBW Baltic 2* started operation in September 2015 and lies directly to the southeast of the FINO2 mast. At FINO3, the *DanTysk* wind farm went into operation in ~~2015~~ December 2014 to the east of the mast. Due to the difficulty in understanding the impacts of the wind turbines on the measurements ~~, the~~ (e.g., times of operation at the onset can vary due to adjustments and testing), the data has not been filtered or corrected for the turbine wakes. However, the presence of the wind farm can impact the evaluation of the model results and should be kept in mind.

All measurement data ~~had~~ is 10 min mean values. ~~The measurement data~~ It was filtered to one period per hour, using the period that was closest to 00 minutes~~, giving a maximum sample size of 8760. The sample size in Table 1 represents the number of samples when all levels used for the wind profile evaluation, indicated in bold, are available. The~~ . The additional filtering for mast flow distortion process removes a small number of samples. For the time-series evaluation~~(e.g. correlation and RMSE)~~, we used measurements from the level that was closest to 100 m above mean sea level and had a good data availability (underlined values, Table 1). The choice of the wind direction height for the model evaluation and data filtering uses the same procedure.

**Table 1.** Tall mast sites and wind measurement (WS: wind speed, WD: wind direction) heights available at each site. The values indicated in bold are used in the evaluation of wind profiles; the values underlined are those used in the evaluation of temporal variability. The centre of the sector where the mast can cause flow distortion and influence the wind speed is also listed. The last three columns show the data availability for 2015 for profiles, time-series and wind direction,respectively.

| site | type | measurement heights (m AGL/AMSL) | flow distortion centre (°) | data availability (%) | | |
|---|---|---|---|---|---|---|
| | | | | profile | time series | wind dir |
| FINO1[a] | Offshore | WS: 32.8, **40.3**, **50.3**, **60.3**, **70.3**, **80.3**, **90.3**, 101.2 | 315 | 78.9 | 82.2 | 88.3 |
| | | WD: 40.7, 50.3, 60.7, 70.3, 80.7, 90.3 | | | | |
| FINO2[a] | Offshore | WD: 32.4, **42.4**, **52.4**, **62.4**, **72.4**, **82.4**, **92.4**, 102.5 | 15 | 77.7 | 82.5 | 97.8 |
| | | WD: 31.8, 51.8, 71.8, 91.8 | | | | |
| FINO3[a] | Offshore | WS: 30.5, 40.5, **50.5**, 60.5, **70.5**, 80.5, **90.5**, 100, 106 | – | 97.3 | 97.3 | 97.6 |
| | | WD: 60.5, 100.5 | | | | |
| IJmuiden[b] | Offshore | WS: **27**, **58**, **89**, **115**, **140**, **165**, **190**, **215**, **240**, **265** | – | 81.5 | 98.5 | 88.5 |
| | | WD: 27, 58, 89, 115, 140, 165, 190, 215, 240, 265 | | | | |
| Høvsøre[c] | Coastal | WS: 10, **40**, **60**, **80**, **100**, **116.5** | 0 | 96.6 | 96.8 | 96.9 |
| | | WD: 10, 60, 100 | | | | |
| Risø[d] | Land | WS: **44.2**, **76.6**, **94**, **118**, **125.2** | 225 | 90.9 | 94.1 | 100. |
| | | WD: 76.5, 94. | | | | |
| Østerild[e] | Land | WS: 10, **40**, **70**, **106**, **140**, **178**, **210**, **244** | 0 | 78.3 | 78.3 | 78.6 |
| | | WD: 40, 244 | | | | |
| Cabauw[f] | Land | WS: 10, **20**, **40**, **80**, **140**, **200** | – | 99.1 | 100. | 100. |
| | | WD: 10, 20, 40, 80, 140, 200 | | | | |

[a] https://www.fino-offshore.de/en/, [b] Kalverla et al. (2017), [c] Peña et al. (2015), [d] http://rodeo.dtu.dk/rodeo/ProjectOverview.aspx?&Project=5&Rnd=674271, [e] Peña (2019)
[f] http://www.cesar-database.nl

## 4 Model evaluation metrics

~~There are many ways of comparing two time series, and the best mathematical tool for such a comparison depends on the relationship being compared.~~ The main goal of the NEWA ~~project~~ sensitivity study was the evaluation of the wind climate, which is usually understood as the probability distribution of wind speed and direction at a specific point. ~~The~~ Thus, we used several metrics to evaluate the accuracy of the model simulations when compared to tall mast observations tailored to this purpose.

We calculate the temporal mean of each modelled distribution, $\overline{u}_m$, and the observed distribution, $\overline{u}_o$, ~~was calculated for identical~~ for identical time periods. The bias herein is defined as difference between the two means, $\overline{u}_m - \overline{u}_o$. If the bias is positive, the model overestimates the observed wind speed.

~~In some applications the temporal accuracy could be important. Although not our primary focus, we calculate time series metrics to gauge the overall quality of the results. If the relative performance of in time-dependent metrics would be significantly different from performance climate wise, that would indicate deep problems in models and would complicate the decisions. However, broadly speaking, that is not the case in our results. The information about temporal co-variability is provided herein by the Pearson correlation coefficient, $r$. The root mean square error (RMSE) is used to provide an estimate of systematic biases in model skill (von Storch and Zwiers, 1999). The RMSE is calculated over all $i$ time steps, with $u_o^i$ and $u_m^i$ being the $i$-th modelled and observed values in the time series of length $n$. The RMSE can be calculated as:~~

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(u_m^i - u_o^i)^2}.$$

~~In the context of this study, a large ensemble of WRF model setups will be compared against the observations at a number of sites. As stated above, one of the ensemble members was designated to be the baseline or "base". The aim is therefore to evaluate if a certain model set up, from the pool, performs better than the baseline configuration. This can be described for a generalised error statistic $m$ defined for the baseline as $m_B$ and $m_{M_j}$ for the $j$-th ensemble member. If the statistic $m$ has the property that the ideal result is $m = 0$ and $m > 0$ means degraded results with respect to the baseline case, then for negative values of the ratio $(m_{M_j} - m_B)/m_B$, the $j$-th member performs better than the baseline and the opposite is also true. Thus, the value of this ratio is the relative improvement or worsening, in percent, compared against the baseline.~~

~~When assessing the change in bias between the $j$-th ensemble member and the baseline, we use the difference in the absolute relative bias, $|(\overline{u}_j - \overline{u}_o)/\overline{u}_o| - |(\overline{u}_B - \overline{u}_o)/\overline{u}_o|$, where the overbar denotes the temporal mean. If the relative bias of the $j$-th ensemble is closer to zero than that of the base, then the $j$-th ensemble member is closer to the observations than the baseline case and the difference of the absolute biases will be negative.~~

~~While the bias is~~ The bias is a popular error statistic for comparing ~~two distributions of wind speed , it suffers from the fact that two distributions can have the same means while having completely different shapes (see Fig. 2 and discussion below). The~~ the wind speed distributions between observations and model-simulated fields. However, the shape of the wind speed distribution ~~plays a large role~~ is more important in wind energy applications, ~~since wind power~~ because the wind power density is proportional to the cube of the wind speed~~; this results in~~. Thus, small changes in the wind speed distribution ~~being~~ are amplified when converted to power.

~~Therefore, we also applied~~ Accordingly, we introduce the Earth Mover's ~~distance (EMD, Rubner et al., 2000) metric, which~~ Distance to evaluate the differences in the shape of two frequency distribution. The EMD, also known as the first Wasserstein distance, is popular in image processing ~~, to evaluate the shape of the distribution. This metric can be understood~~ (EMD; Rubner et al., 2000) The EMD can be interpreted as the amount of physical work needed to move a pile of soil in the shape of one distribution to that of another distribution. ~~The EMD~~ More discussion about the EMD properties can be found in Lupu et al. (2017). For one-dimensional distributions the EMD is equivalent to the area between two cumulative distribution functions, and, this interpretation with slight modifications, can be applied also to circular variables (Rabin et al., 2008). The EMD ~~calculation~~ was calculated using the Pyemd package (Pele and Werman, 2008).

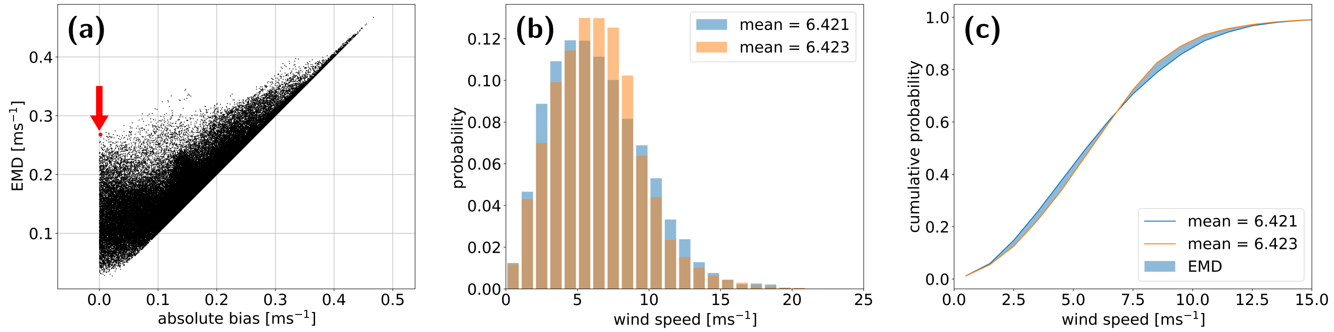
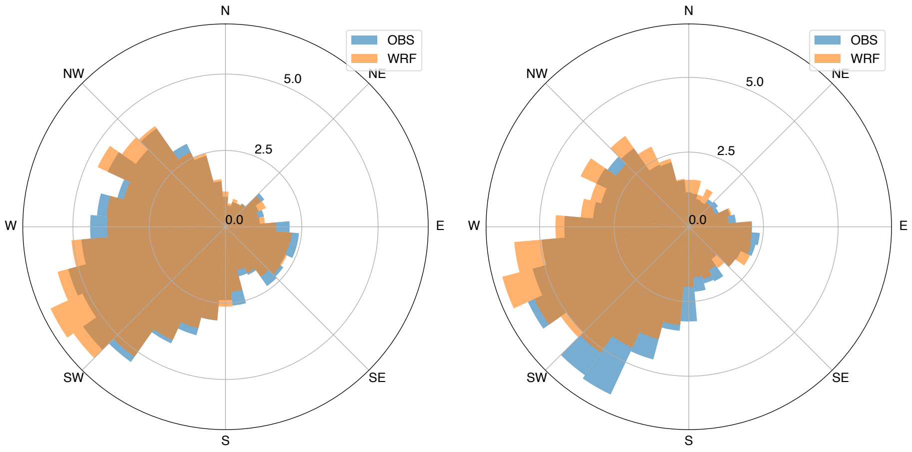


**Figure 2.** (a) Relationship between EMD values and absolute value of bias between the wind speed in two WRF model setups for all grid-points in the domain. The red dot represents an example where the EMD and the absolute value of bias are different. (b) Wind speed distributions of the two ensemble members corresponding to the red dot. (c) Cumulative probability distribution together with the EMD value calculated as the area between them. Distributions are the same as in panel (b).

Figure 2 illustrates the differences between the EMD and the absolute value of the bias. ~~The left panel~~ Figure 2a shows the relationship between the EMD and the absolute value of bias for two WRF model simulations for all points in the domain~~, with each dot in the plot representing one grid point. The right panel of~~ . Figure 2b shows modelled wind speed distributions for two separate grid points~~to highlight a~~ . This highlights the case where two differently shaped distributions can have the same mean~~, and using the EMD metric can identify the differences in such distributions. The values for EMD and difference in means are similar when both of~~ . The EMD helps clarify such occurrences. For one-dimensional distributions the EMD can be calculated as the area between the cumulative distribution curves, as illustrated in Figure 2c.



**(a) FINO3, CEMD=2.18°** **(b) FINO1, CEMD=5.26°**

**Figure 3.** Wind direction distributions for (a) FINO3 and (b) FINO1 and using measurements (OBS) and model simulated (WRF) data.

The circular EMD (CEMD; Rabin et al., 2008) extends the EMD concept to one-dimensional circular histograms, such as the frequency distribution of wind directions. Two examples of the value of the CEMD are given in Fig. 3 for FINO3 and FINO1. At FINO3, the observed and simulated wind direction distributions are very similar and the value of CEMD is 2.18°, mainly due to differences in frequency within the same sector. A higher value of CEMD (=5.28°) is obtained when a rotation in the wind direction is found between the ~~numbers are large.Therefore, using EMD instead of bias provides the greatest value when it is necessary to distinguish between two differently-shaped distributions, which might have the same means. The EMD metric has the same units as the variable being compared~~observed and simulated wind directions. The CEMD is used throughout the paper, and is to our knowledge the first time that it is used for evaluating wind directions in meteorology or climate science.

The information about temporal co-variability is provided herein by root mean square error (RMSE), which estimates of systematic biases in model skill (von Storch and Zwiers, 1999). The RMSE is calculated over all $i$ time steps, from

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(u_m^i - u_o^i)^2}, \tag{1}$$

with $u_o^i$ and $u_m^i$ being the $i$-th modelled and observed values in the time series of length $n$. ~~When comparing the performance of the~~

In this study, we compared a large ensemble of WRF model setups against the observations at several sites. One of the ensemble members was designated to be the baseline or “BASE”. The aim is to evaluate if a certain model set up from the pool performs better or worse than the baseline configuration. For this purpose we define a general skill score (von Storch and Zwiers, 1999):

$$\mathrm{SS} = 1 - M_j/M_B, \tag{2}$$

where $M_j$ is the value of the metric for the $j$-th ensemble member ~~to~~ and $M_B$ is that of the baseline~~, we use the ratio (EMD$_j$ − EMD$_B$)/EMD$_B$; if this difference is negative,~~ . The metric $M$ can be the absolute value of the bias, the ~~time series of the ensemble verifies better than the baseline against the observations.~~ RMSE, EMD, or CEMD. If $\mathrm{SS} > 0$ the ensemble member $j$ “improves” the metric with respect to the baseline case, if $\mathrm{SS} < 0$ it “worsens” it. A value of $\mathrm{SS} = 1$ means that the new simulation is perfect. The SS is easily understood and is applicable to all our evaluation metrics. However, when the BASE simulation evaluates extremely well against observations (e.g. when the bias is close to zero), the skill score can become very large. Therefore, the SS is a useful quantity for the RMSE, which is rarely close to zero, but can be misleading when used for the absolute bias or the EMD, indicating large improvements when the differences in metrics themselves are small. Accordingly, we suggest using both SS and the original metric when interpreting the results.

## 5 Sensitivity analysis of WRF simulations

In this section, the results from the different sensitivity experiments are presented and discussed. These are grouped into ~~five~~ six subsections: Section 5.1 presents the results from five different domains for a small number of experiments to see how the results differ depending on the simulated region; Section 5.2 evaluates the results from one of these domains against observations; Section 5.3 highlights the impact of the WRF version on the wind speed results by investigating four different versions of the WRF model; Section 5.4 presents the results from a 25 simulation sensitivity study addressing the impact of the SL, PBL and LSM schemes, including changes to the surface roughness length; Section 5.5 documents the impact of other parameterizations and forcing data; and finally Section 5.6 focuses on the impact of the size of the domain. It should be noted that these sections do not necessarily follow the chronological order that they were performed in the NEWA project, but provide a logical progression of the decisions taken during the project.

### 5.1 Sensitivity to geographical domain

In the initial stage of the evaluation of the WRF model setup, we designed four numerical experiments (top four simulations in Table 2) over five different regions in Europe (Fig. 4), mostly located near countries represented in the NEWA team. The ~~aim was to~~ experiments explore the impact of using different PBL schemes~~either~~: the MYNN (Mellor and Yamada, 1982) or ~~YSU (Hong et al., 2004),~~ the YSU (Hong et al., 2004) and the effect of ~~using~~ different initialisation strategy ~~: either using shorter or longer~~ (short, S1, or weekly, W1, simulation length and exclude or include nudging~~. This series of numerical experiments had the main objective~~). The YSU and MYNN schemes are non-local-K mixing and TKE 1.5 order closure models, respectively. They were chosen because they are popular among the WRF model users and have shown good skill in previous studies (Draxl et al., 2014; Hahmann et al., 2015). The main objective was to clarify whether the sensitivity of the mean wind speed to these changes is similar in different geographic regions or ~~if~~ there were regional differences. All other settings were ~~left~~ fixed. The simulations were carried out with the WRF model version 3.6.1, released in August 2014. The basic WRF model setup includes ~~the use of~~ using the ERA-Interim (Dee et al., 2011) for initial and boundary conditions, NCEP optimal interpolation sea surface temperature (SST, Reynolds et al., 2007) and 61 vertical levels. Other details are given in Table A1.

**Table 2.** Acronyms and relevant set up parameters of the WRF model sensitivity experiments oriented to address the influence of the geographic domain and WRF model version. The mod MYNN scheme is described in section 5.3.

| experiment | WRF model ~~scheme~~ version | PBL scheme | run length [d] | spin-up length [h] | nudging |
|---|---|---|---|---|---|
| MYNL61S1 | 3.6.1 | MYNN | 1.5 | 12 | no |
| MYNL61W1 | 3.6.1 | MYNN | 8 | 24 | yes |
| YSUL61S1 | 3.6.1 | YSU | 1.5 | 12 | no |
| YSUL61W1 | 3.6.1 | YSU | 8 | 24 | yes |
| MYNL61W1_V381 | 3.8.1 | MYNN | 8 | 24 | yes |
| MYNL61W1_V381_MOD | 3.8.1 | mod MYNN | 8 | 24 | yes |

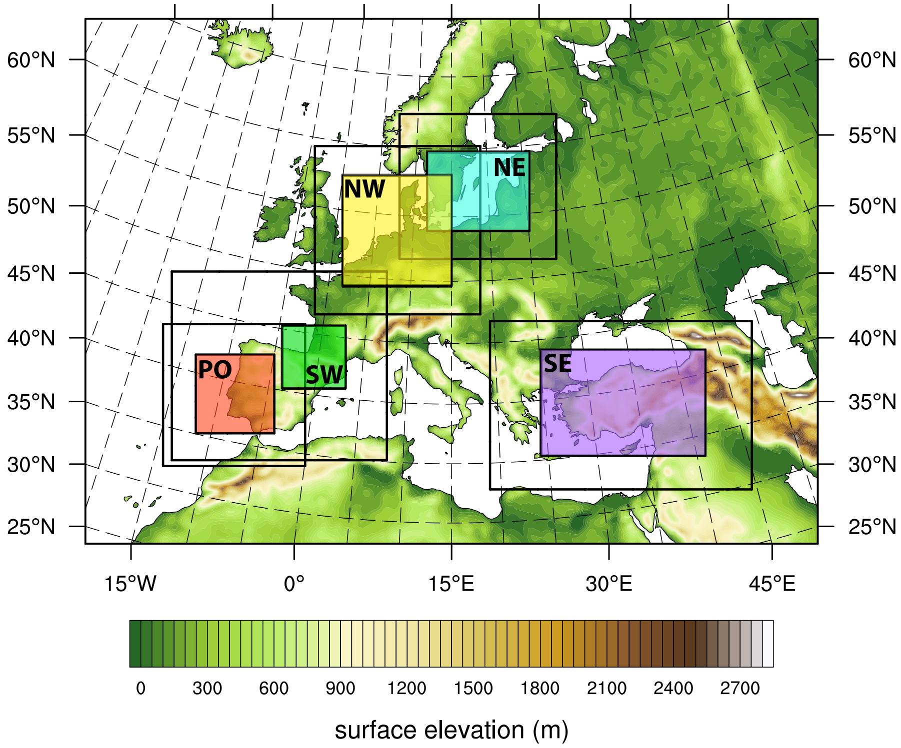


**Figure 4.** The location of the five inner domains (D3; NW, NE, SE, SW, and ~~PD~~ PO in coloured boxes) used in the geographic similarity experiments. The surface elevation of the WRF model outer domain (D1) is shown with colour; the black lines show the extent of D2. ~~The black dots show the locations of NEWA experimental sites (Mann et al., 2017).~~ All inner domains share the same outer domain (D1), which corresponds to the area of the base map.

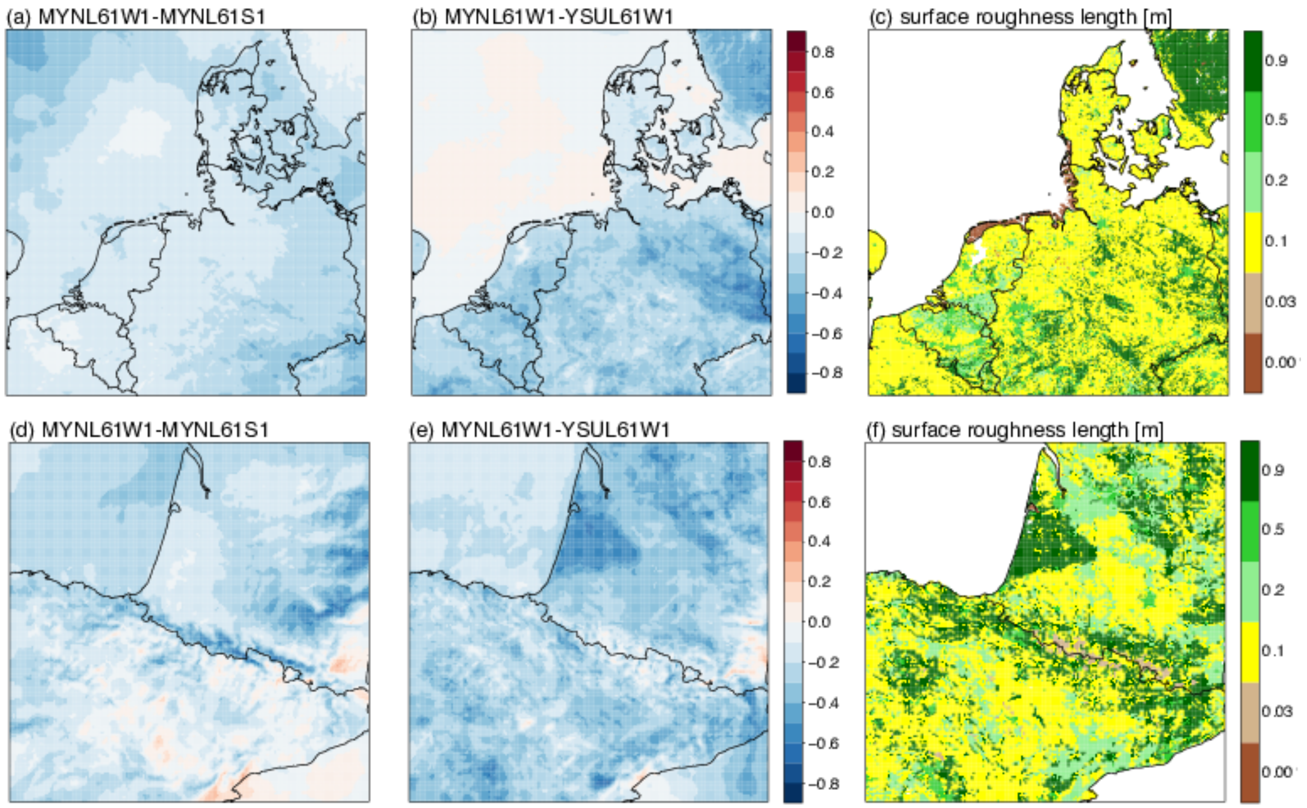


**Figure 5.** Annual mean difference in wind speed [$\mathrm{m\,s^{-1}}$] between the simulations with: different initialisation strategy for the MYNN simulations (W1 minus S1; a and d), different PBL scheme (MYNN minus YSU; b and e), and the surface roughness length [m] used in the simulations (c and f). (a), (b) and (c) are for the NW domain, (d), (e) and (f) are for the SW domain shown in Fig. 4.

The analysis of the experiments ~~in Table~~presented in Fig. ~~2 showed that the largest differences~~ 5 show that on average the differences in annual mean wind are small and those that arise from the choice of PBL scheme ~~, as shown in~~ are larger and more extensive than those from the initialisation strategy. The largest differences in wind speed that arise from the initialisation strategy (Fig. 5~~for the NW and SE domains.The left side of~~a and Fig. 5d) coincide with areas of elevated terrain in the western French Alps and the Pyrenees in the ~~figure shows the differences in annual mean wind speed at 100~~SW domain and the north-west corner of the NW domain. In these areas the daily runs (S1) have on average stronger winds that the weekly nudged simulations (W1). The largest differences in wind speed that arise from the choice of PBL scheme (Fig. ~~m between simulations using the MYNN and YSU PBL schemes; the plots on the right side show the surface roughness length for the two domains. The results show that the regions with the largest differences~~ 5b and Fig. 5e) coincide with the regions with particularly large surface roughness length, namely forests in southern Sweden and south-western France. There, the experiment using the MYNN scheme provides wind speeds that are on average more than $0.5\,\mathrm{m\,s^{-1}}$ lower than in the experiment using the YSU scheme. Over the sea, no significant difference is seen in the NW domain, and only slight differences (less than 0.3 m s$^{-1}$) exist above the Atlantic and Mediterranean (SW domain).

~~Annual mean difference in wind speed m s$^{-1}$between the MYNL61W1 and YSUL61W1 simulations for the (a) NW and (b) SW domains. Surface roughness length mfor the (c) NW and (d) SW domains.~~

Similar analysis was carried out for the other three domains in Fig. 4. All five domains show the same pattern of higher wind speed at 100 m for simulations using the YSU scheme over land, with the largest differences occurring over rougher terrain (e.g. forests). Over water, the differences are much smaller~~, but~~. The winds simulated using the YSU scheme are slightly higher than those simulated using the MYNN scheme, but only in regions dominated by unstable stratification (e.g. the Atlantic Ocean, Mediterranean and Black Seas).

~~In the~~ In the nudged weekly simulations, winds speeds are overall reduced from those in the daily simulations, and to a higher magnitude over higher terrain and closer to the edge of the domain. Similar results were obtained in Hahmann et al. (2015) when the discarded spin-up time of the short simulations was varied. We hypothesise that as the model integration progresses, a balance is reached between the large-scale flow, the model physical parameterisations and the surface forcing supplied by the terrain elevation and surface roughness. This process takes some time and results in lower wind speeds in the longer simulations. The effect is different on the northwest corner of the NW domain, which is the dominant inflow boundary, probably resulting from the nudging in the outer domain that is absent in the short runs.

### 5.2 Evaluation against observations in the NW domain

The evaluation of the mean wind speeds from the four WRF V3.6.1 sensitivity experiments against the mast measurements ~~(~~is shown in Fig. 6 ~~), the differences~~ and Fig. 7. The differences in mean wind speed between the various experiments are small~~, but overall the mean winds~~. The mean wind profile from the MYNL61W1 ~~run are~~ simulation is closest to the observed ~~value~~ profile at nearly all the sites, except at ~~FINO2. The mean statistics~~ the offshore sites where the MYNL61W1 and YSUL61W1 are nearly indistinguishable from each other. The evaluation metrics of the wind speed and direction for the experiments and sites are presented in Fig. 7. This confirms that the wind speeds from the MYNL61W1 ~~run~~ simulation have: the lowest biases at Cabauw, Høvsøre, Østerild, and Risø; the ~~highest correlation at all sites; and the~~ lowest RMSE at all sites, except for FINO1 and IJmuiden, where the results from MYNL61W1 and YSUL61W1 are virtually the same. The EMD shows lowest values for the weekly nudged simulations compared to the short runs and for the coastal and land sites the lowest EMD is for the MYNL61W1 simulation. Similar conclusions apply to the CEMD of the wind direction, where the YSUL61W1 performs the best at six of the eight sites.

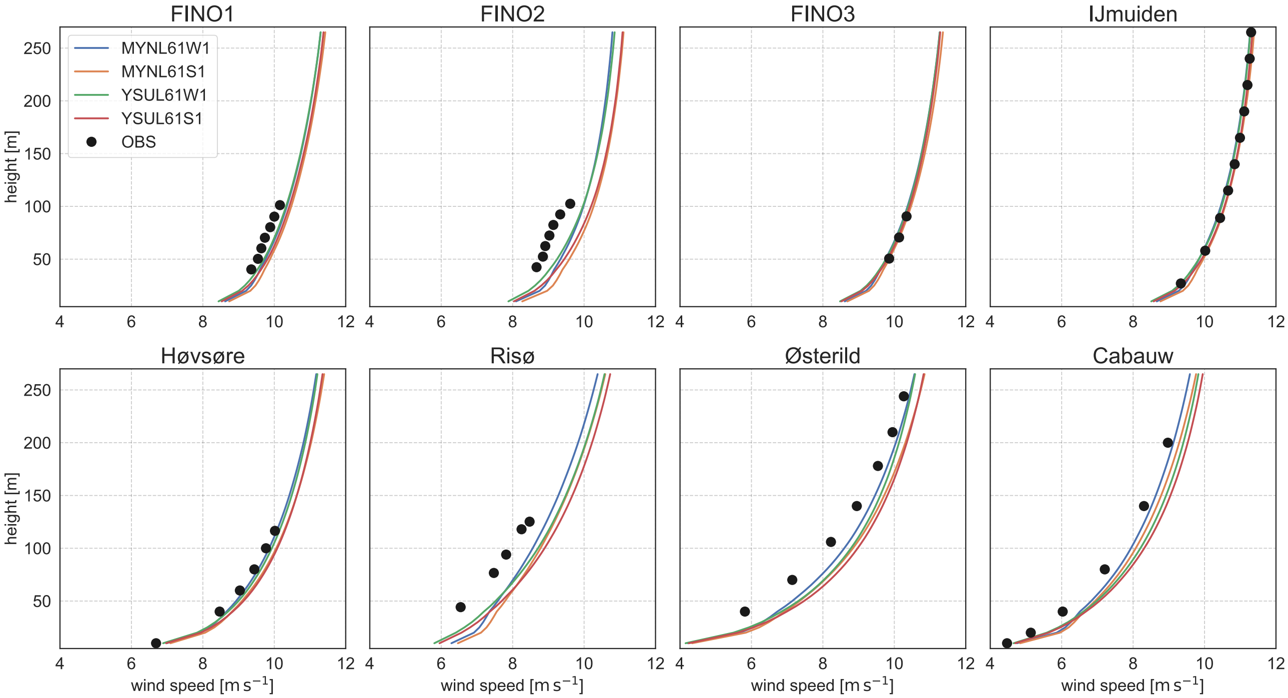


**Figure 6.** Comparison of the observed mean wind speed [m s$^{-1}$] as a function of height for the eight sites and the simulated mean wind speed from the four sensitivity WRF V3.6.1 runs in Table 2.

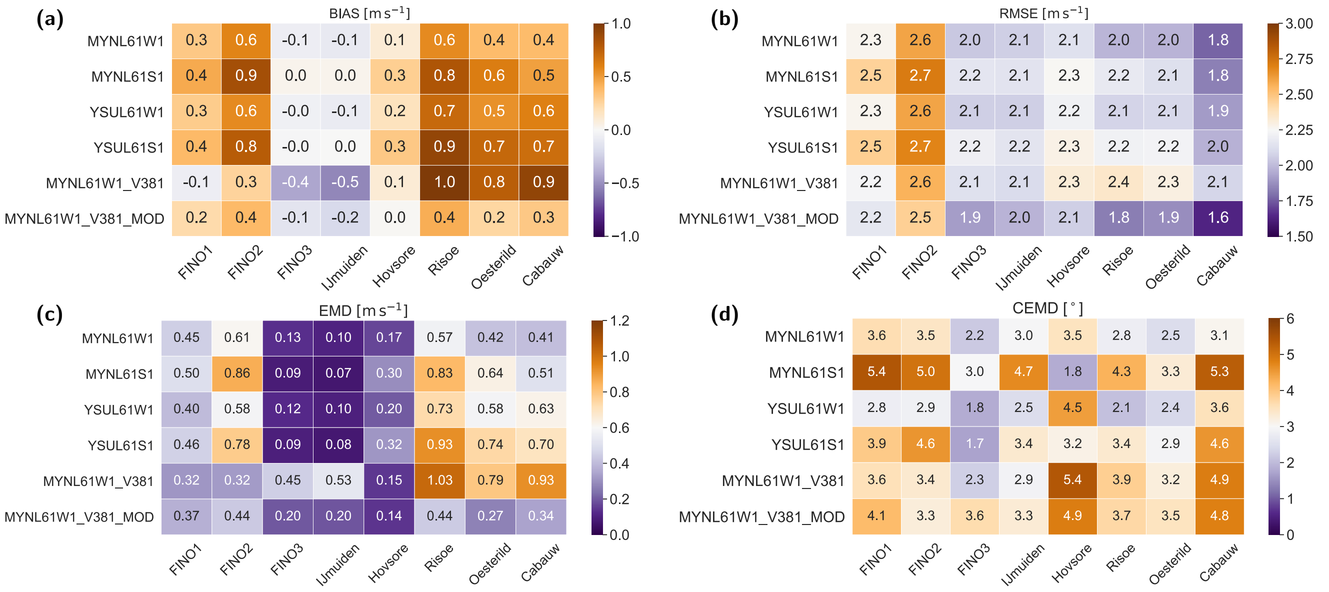


**Figure 7.** Evaluation metrics: (a) bias [m s$^{-1}$], (b) RMSE [m s$^{-1}$], (c) EMD [m s$^{-1}$] for the simulated wind speed and (d) CEMD [°] for simulated wind direction for the eight sites and underlined heights in Table 1 and the sensitivity experiments in Table 2

In summary, the weekly nudged simulations for both MYNN and YSU schemes result in lower biases and ~~higher correlations~~ lower RMSE at all sites. ~~Also, because~~ Because the effect on the wind speed ~~between the two~~ when one of those PBL schemes is replaced with another is nearly insensitive to the location of the domain (i.e. North versus South) as shown in ~~Fig.5, this should be~~ Section 5.1, we continue under the assumption that it is valid for other regions in Europe, except for regions with more complex terrain, which have not been evaluated. The weekly nudged simulation setup was chosen for the remainder of the NEWA sensitivity simulations. The ~~use of~~ setup and constants used in the nudging will be re-evaluated in the sensitivity experiments in Section 5.5.

## 5.3 Sensitivity to properties of the ~~WRF model version~~MYNN scheme

~~The WRF model version was also tested~~ We also tested the version of the WRF model used during the sensitivity analysis to evaluate whether changing the version implied any difference with respect to the baseline configuration described above. At the time of the development of this work, the latest version was WRF version 3.9.1, released in August 2017. The simulations using versions WRFV3.8.1 and WRFV3.9.1 ~~for~~ use the same NW domain and the same model setup as in the MYNL61W1 ~~were carried out~~simulations (Table 2). The results for WRFV3.6.1 and WRFV3.8.1 are presented in Fig. 7 and Fig. 8; the results for WRFV3.9.1 (not shown) are ~~almost~~ nearly identical to those from WRFV3.8.1.

The wind speed from simulations using WRFV3.8.1 presented an increased bias compared to observations for all sites and most levels, except for FINO1 and FINO2, which suffer from flow distortion and wind farm effects. The increased bias was traced back to changes in two important equations in the MYNN SL and PBL scheme, which became defaults in WRF Version

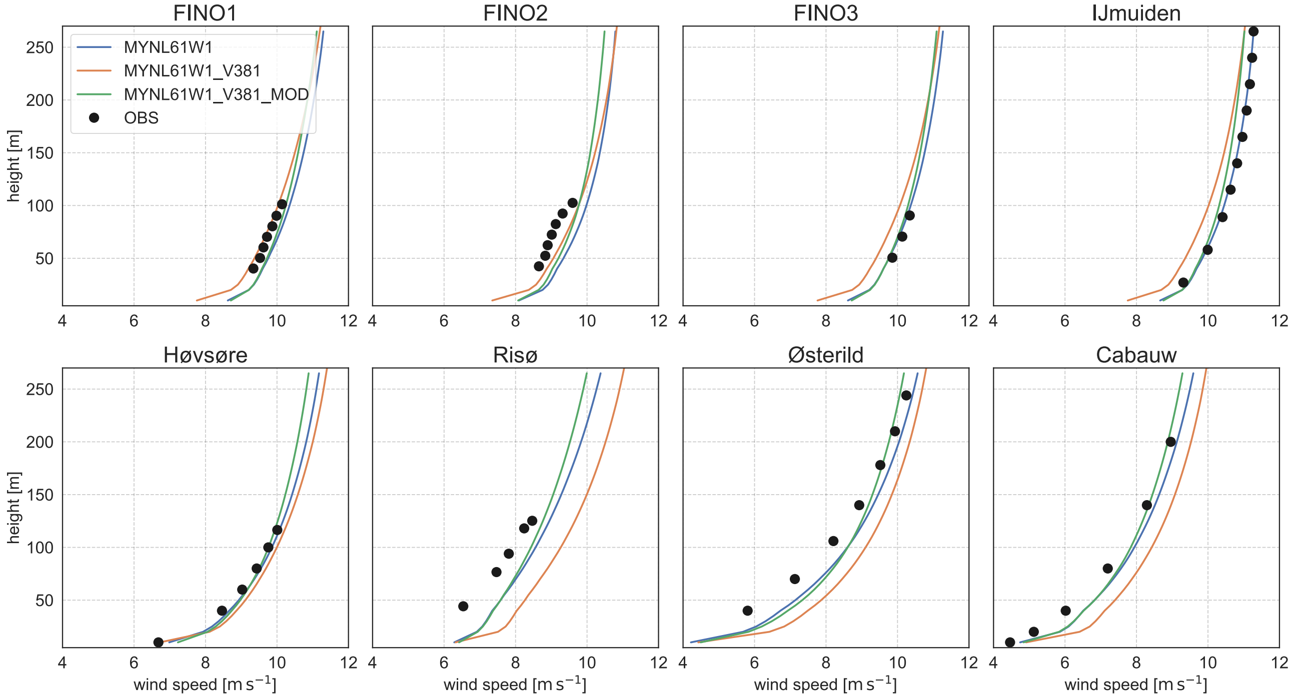


**Figure 8.** Comparison of the mean observed wind speed [m s$^{-1}$] as a function of height for the eight sites and the mean wind speed in the simulations using WRFV3.6.1, WRFV3.8.1 and the modified version of WRFV3.8.1_MOD described in Table 2.

3.7. The first is the scalar roughness length over water, which was changed from the formulation in Fairall et al. (2003) to that in Edson et al. (2013), thus affecting the wind speed over the ocean. The second is a change in the definition of the mixing length (Olson et al., 2016). Both of these options could be customised and set as in the baseline configuration. Results from such a setup are labelled "WRFV3.8.1_MOD" in Fig. 8. Some of the previous characteristics of the profile are restored after these changes, and the simulation using WRFV3.8.1 with modifications improved the RMSE and EMD for all sites ~~at ∼~~(fig. ~~100 m (not shown).~~7. These changes are consistent with those found by Yang et al. (2017) for the MYNN scheme. Although above ∼ 100 m, the simulation using the modified WRFV3.8.1 gives lower mean wind speeds than WRFV3.6.1 at all sites, we consider nonetheless that this ~~differences~~ difference is less relevant and based on the improvements of the scheme "WRFV3.8.1_MOD", this set up was selected as the baseline, named "BASE" hereon. The WRF model setup of this baseline is summarised in Table A1. Unless explicitly labelled otherwise, when referring to the MYNN option in the remainder of this work we mean the modified version of the scheme. The unmodified MYNN PBL will be referred to as MYNN*.

### 5.4 Effect of surface and planetary boundary layer and land surface model

The first series of sensitivity studies tested the sensitivity of the near surface wind to various combinations of LSM, PBL, SL schemes and the specification of surface roughness length. The ~~schemes~~ combinations tested are listed in Table 3. All the simulations in this and most of those in Section 5.5 use the ERA5 reanalysis (Hersbach et al., 2020) as the source of initial and boundary conditions.

The large number of schemes and their potential combinations led to a large number of possible combinations. In this work a total of 25 different combinations were tested, listed in Table 3, including changes in parameters in the schemes themselves. We also included the simulation using the ~~unmodified MYNN PBL (labelled~~ MYNN* ~~, see Section 5.3)~~ scheme. Our original table of sensitivity experiments contained many other LSM/PBL/SL combinations, but some of these suffered from diverse technical issues and did not complete the runs. In some cases, small adjustments were needed~~, that is the fractional sea-ice option had to be turned off when~~. For example, in the experiments using the MM5 or MO surface layer schemes the fractional sea-ice option had to be turned off.

An important aspect of the LSM/PBL/SL sensitivity studies was the use of an alternative look-up table for the surface roughness length as a function of land-use class. A custom NEWA lookup table was created since many values used in the WRF-distributed tables do not match the aerodynamic characteristics of European vegetation, especially over forests (Hahmann et al., 2015; Floors et al., 2018a). The new lookup table was created by polling wind energy resource assessment experts from the NEWA consortium. Both the new and old values of surface roughness length for each roughness class are shown in Table 4. Some of the larger changes include, “Herbaceous Wetland”, which has an original value of $z_0 = 0.20\,\mathrm{m}$, but in the NEWA ~~regional~~ region represents the tidal zone in coastal Holland, Germany, and Denmark, which is much smoother (Wohlfart et al., 2018), and thus was changed to $z_0 = 0.001\,\mathrm{m}$. The forest classes were also significantly changed, with the $z_0 = 0.50\,\mathrm{m}$ value in the default table being changed to $z_0 = 0.90\,\mathrm{m}$, which is more representative of forests in, for example Sweden (Dellwik et al., 2014). The new roughness values should be considered only as estimates and as such there might be some limitations in the representation of the roughness length, they are nevertheless much more realistic than the default ones. The experiments using the standard vegetation tables are labelled WRF vegetation in Table 3.

All NEWA setups use a constant value of surface roughness and have no annual cycle, except for two of the setups (ANNZ0 and ANNZ0N) that have annual cycle according to the default WRF vegetation table (except for the tidal zone). In WRF, the seasonality of the surface roughness length is controlled by the value of the green vegetation fraction (Refslund et al., 2014) and applies mostly for cropland classes in Table 4. The annual cycle of green vegetation fraction does not change from year to year and is spatially inconsistent with the ESA-CCI land-use dataset used in the NEWA simulations. Therefore, because of inherent uncertainties, in NEWA we have chosen to use a single constant value of roughness for land-use and land cover class. Also, since the WRF wind climatologies will be further downscaled, as described in Dörenkämper et al. (2020), using a constant value of surface roughness facilitates the process. Another roughness-related experiment that was included, AGGZ0, uses the sub-tiling option for NOAH (Li et al., 2013), with the NEWA vegetation table. The sub-tiling option generates more realistic values of surface roughness length in areas of mixed vegetation, which could reduce the biases in wind speed (Santos-Alamillos et al., 2015).

The vertical profiles of mean wind speed for all the LSM/PBL/SL sensitivity experiments for four of the eight evaluation sites are shown in Fig. 9 as an example. ~~It is difficult to distinguish between the results of the various setups, but generally, the setups using the MYNN PBL scheme (in blue) tend to have lower wind speeds, which are often closer to the observed values~~

**Table 3.** Overview of the ensemble of WRF model simulations varying LSM/PBL/SL scheme and vegetation lookup table carried out for the NW domain. The names of the schemes are: LSM: NOAH (Tewari et al., 2004), RUC (Benjamin et al., 2004), PX (Noilhan and Planton, 1989), SLAB (Dudhia, 1996), NOAH-MP (Niu et al., 2011); PBL/SL: YSU (Hong et al., 2006), MYJ (Mellor and Yamada, 1982), MYNN (Nakanishi and Niino, 2006), modified MYNN see Section 5.3), ACM2 (Pleim, 2007), MM5 (Jiménez et al., 2012), M-O surface layer (Janjic and Zavisa, 1994). The number in parenthesis is the respective namelist value in the WRF model configuration file. The "simulation code" is explained in Appendix A2.

| run name | simulation code | LSM (#) | PBL (#) | SL (#) | veg table |
|---|---|---|---|---|---|
| BASE | EES81_2551040004 | NOAH (2) | MYNN (5) | MYNN (5) | NEWA |
| MYNN* | EES81_2550040004 | NOAH (2) | ~~MYNN*~~MYNN[a] (5) | ~~MYNN*~~MYNN[a] (5) | NEWA |
| MM5 | EES81_2511040004 | NOAH (2) | MYNN (5) | MM5 (1) | NEWA |
| MO | EES81_2521040004 | NOAH (2) | MYNN (5) | M-O (2) | NEWA |
| MYJ-MO | EES81_2220040004 | NOAH (2) | MYJ (2) | M-O (2) | NEWA |
| YSU-MM5 | EES81_2110040004 | NOAH (2) | YSU (1) | MM5 (1) | NEWA |
| RUC | EES81_3551040004 | RUC (3) | MYNN (5) | MYNN (5) | NEWA |
| RUC-WRF | EES81_3551040004_A | RUC (3) | MYNN (5) | MYNN (5) | WRF |
| RUC-MO | EES81_3521040004 | RUC (3) | MYNN (5) | M-O (2) | NEWA |
| RUC-YSU-MM5 | EES81_3110040004 | RUC (3) | YSU (1) | MM5 (1) | NEWA |
| RUC-ACM2-PX | EES81_3770040004 | RUC (3) | ACM2 (7) | ACM2 (7) | NEWA |
| PX-ACM2-PX | EES81_7770040004 | PX (7) | ACM2 (7) | ACM2 (7) | NEWA |
| PX-ACM2-MM5 | EES81_7710040004 | PX (7) | ACM2 (7) | MM5 (1) | NEWA |
| SLAB | EES81_1551040004 | SLAB (1) | MYNN (5) | MYNN (5) | NEWA |
| SLAB-MYJ-MO | EES81_1220040004 | SLAB (1) | MYJ (2) | M-O (2) | NEWA |
| SLAB-YSU-MM5 | EES81_1110040004 | SLAB (1) | YSU (1) | MM5 (1) | NEWA |
| SLAB-ACM2-PX | EES81_1770040004 | SLAB (1) | ACM2 (7) | ACM2 (7) | NEWA |
| NOAHMP | EES81_4550040004 | NOAH-MP (4) | MYNN (5) | MYNN (5) | NEWA |
| NOAHMP2[b] | EES81_4550040004_B | NOAH-MP (4) | MYNN (5) | MYNN (5) | NEWA |
| NOAHMP-WRF | EES81_4550040004_A | NOAH-MP (4) | MYNN (5) | MYNN (5) | WRF |
| NOAHMP-MYJ-MO | EES81_4220040004 | NOAH-MP (4) | MYJ (2) | M-O (2) | NEWA |
| NOAHMP-YSU-MM5 | EES81_4110040004 | NOAH-MP (4) | YSU (1) | MM5 (1) | NEWA |
| ANNZ0[c] | EES82_2551040004 | NOAH (2) | MYNN (5) | MYNN (5) | WRF[d] |
| ANNZ0N[c] | EES82_2551040004_A | NOAH (2) | MYNN (5) | MYNN (5) | WRF |
| AGGZ0[c] | EES83_2551040004 | NOAH (2) | MYNN (5) | MYNN (5) | NEWA |

[a] is the unmodified MYNN scheme, [b] uses opt_sfc = 2 in the NOAH-MP scheme. [c] differs in surface roughness, see text for details. [d] WRF table, but low roughness for tidal zone.

~~over the sea.~~ The spread in the wind speed among the simulations excluding outliers generally increases with height reaching around 1 m s$^{-1}$ at 100 m~~over the sea and around 1.5~~. Over land, for example at Østerild, the observed wind speed profile below

**Table 4.** Vegetation look-up table for the surface roughness length as a function of the USGS land use category (Anderson et al., 1976) in the NEWA and default NCAR WRF model configuration. Only values changed from default are shown.

| USGS type | land-use land cover class | $z_0$ NEWA [m] | $z_0$ WRF orig range [m] |
|---|---|---|---|
| 2 | Dryland Cropland and Pasture | 0.10 | 0.05–0.15 |
| 3 | Irrigated Cropland and Pasture | 0.10 | 0.02–0.10 |
| 4 | Mixed Dryland/Irrigated Cropland and Pasture | 0.10 | 0.05–0.15 |
| 5 | Cropland/Grassland Mosaic | 0.10 | 0.05–0.14 |
| 7 | Grassland | 0.10 | 0.10–0.12 |
| 8 | Shrubland | 0.12 | 0.01–0.05 |
| 9 | Mixed Shrubland/Grassland | 0.12 | 0.01–0.06 |
| 11 | Deciduous Broadleaf Forest | 0.90 | 0.5 |
| 12 | Deciduous Needleleaf Forest | 0.90 | 0.5 |
| 13 | Evergreen Broadleaf Forest | 0.90 | 0.5 |
| 14 | Evergreen Needleleaf Forest | 0.90 | 0.5 |
| 15 | Mixed Forest | 0.50 | 0.20–0.50 |
| 17 | Tidal zone[a] | 0.001 | 0.20 |

[a] Originally called “Herbaceous Wetland” in the default WRF model vegetation table.

$\sim 150$ m ~~$s^{-1}$ over land~~lies well below the simulated values for all experiments, signifying a lack of representativeness of the surface conditions in the WRF model at this site, since the biases are independent of the LSM/PBL/SL used.

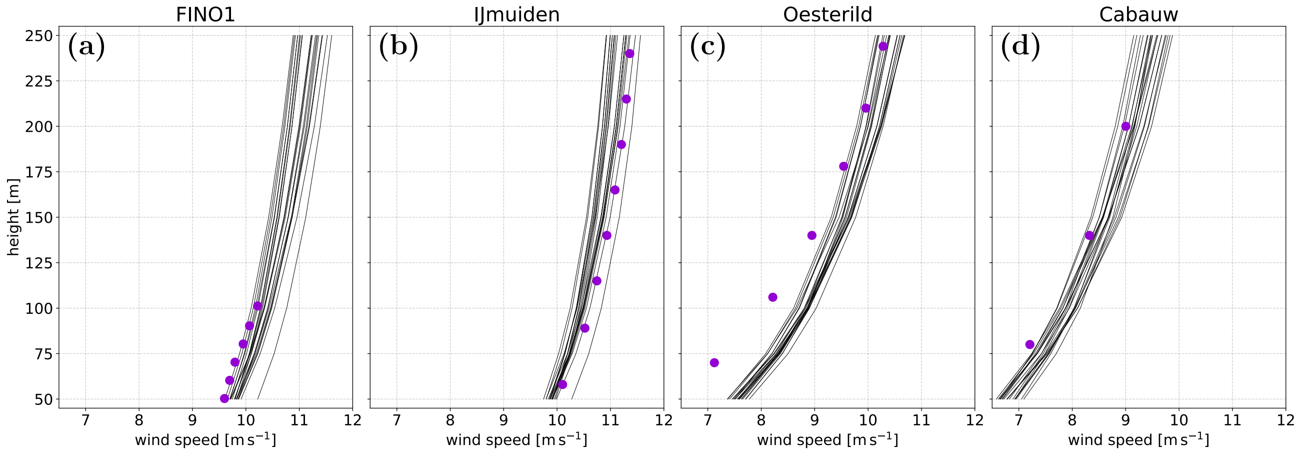


**Figure 9.** Mean wind speed [m s$^{-1}$] simulated by the various experiments in the LSM/SL/PBL ensemble as a function of height for four sites: (a) FINO1, (b) IJmuiden, (c) Østerild and (d) Cabauw. The observed values are shown as purple dots.

To facilitate the intercomparison among the ensemble members, we computed all the evaluation metrics of the wind speed and direction for each simulation (Fig. ~~?? and ??~~10). The ~~left panel of Fig.~~ metrics compare the filtered wind speed observations and wind directions during 2015 at the levels underlined in Table ~~?? shows the relative~~ 1 with the corresponding WRF-simulated time series interpolated to the same height. Figure 10a shows the model bias at all the sites. ~~It shows that the bias shows a certain relation with the site~~For this metric differences among stations are larger than differences between models in a single station, expressed in this figure as consistent colours for each column. ~~Additionally,~~ On average the characteristic of the bias relates most directly to the the quality of the measurements and type of site, ~~i.e.~~ that is slightly negative bias over the sea and positive bias over land. The latter is likely a consequence of deficient representation of the land characteristics around each site~~, since they are independent of the LSM/PBL/SL used~~. Some other general patterns are that the simulations using the MYJ scheme tend to have largest absolute biases, except at FINO3, and that the YSU-MM5-RUC simulation is an outlier, whose results differ from other setups, typically being among the worst of setups for any station~~, which is also evident in the vertical profiles at FINO1 and IJmuiden in Fig. 9.~~.

To better quantify the differences between the simulations, ~~the right panel of~~ Fig. ~~?? shows the absolute difference in the relative~~ 10b shows the skill score (SS) of the bias from the ~~“BASE ”~~ BASE simulation as defined in Section 4. ~~Negative numbers~~ Positive numbers (in purple) show a decrease in absolute bias, which ~~will~~ point to a more accurate simulation. From these values some conclusions can be drawn. First, the differences of the simulations are quite modest, with a maximum ~~of 8.9 % in the MYNN* simulation at Cabauw and~~ SS of 0.9 in the MM5 and MYJ-MO simulation at FINO1 and FINO3 and large up to −~~1.9 %~~ 16.2 in several simulations at ~~various sites~~Høvsøre, which are an artefact of the very low bias in the BASE simulation at this site. Second, no simulation is capable of improving or degrading the bias ~~statistics~~ at all of the sites: the ~~MYJ-MO-SLAB~~ SLAB simulation improves the bias (i.e., SS > 0) at three of the eight sites, while the MO simulation

~~performs better at four~~ improves the bias at five of the eight sites. The ~~unmodified MYNN~~ MYNN* scheme considerably degrades the simulations at ~~six~~ seven of the eight sites. The latter supports our decision to use the modified version of the MYNN scheme as baseline. It is relevant to note that the changes in $z_0$ only cause minor variations in the biases (see members ANNZ0, ANNZ0n, AGGZ0), however the sites are located in regions with vegetation classes that ~~did not change~~ were not changed significantly from the WRF model standard table (Table 4).

~~(a) Biases %and (b) changes in the biases from the BASE simulation |BIAS| − |BIAS_B|, %between the observed and simulated wind speed at the eight sites and the various sensitivity studies in the LSM/SL/PBL ensemble (Table 3).~~

~~(a) EMD relative to the observed wind speed %and (b) relative change in the EMD from the BASE simulation (EMD − EMD_B)/EMD_ %between the observed and simulated wind speed at the eight sites and the various sensitivity studies in the LSM/SL/PBL ensemble (Table 3).~~

~~Fig.~~Figures ~~?? provides~~ 10c and 10d provide further information about the sensitivity tests based on the EMD metric defined in Section 4 to evaluate the shape of the wind speed distributions. As with the bias, the EMD ~~metric~~ shows that the largest differences in total error are ~~related~~ linked to the site location, with the best model performance at Høvsøre, with EMD between ~~(0.6–3.2~~0.1–0.4 ~~%)~~$\mathrm{m\,s^{-1}}$, while at Risø the results fall between ~~(6.8–15.3~~0.4–1.0 ~~%)~~$\mathrm{m\,s^{-1}}$. Particularly interesting is the comparison between ~~the two metrics,~~ bias and EMD, since for most setups and stations the values of EMD and bias are similar, especially when the model results are significantly different from the observations. ~~However, for FINO1, if only the bias was analysed, it could be argued that the base setup represents the observed distribution perfectly. However, the EMD shows that it is not the case and overall performance of the model is comparable to IJMuiden.~~ The EMD allows us to identify cases where the change in model setup improves only the mean value of distribution, as opposed to the similarities of the whole distribution. For instance, if only the biases were analysed, it could be argued that MYJ-MO is ~~significantly~~ better than the ~~base~~ BASE in FINO3 and IJmuiden, while the EMD ~~results~~ show that there is only a modest improvement over the ~~base~~ BASE at IJmuiden and a worse result at FINO3. Similar conclusions can be ~~carried out about other runs using the MYJ-MO PBL scheme~~ drawn for the SLAB-MYJ-MO and YSU-MM5-RUC ~~run. However, as~~ simulations. As with the bias, no simulation improves the EMD for all sites, and very few simulations improve the EMD ~~metric~~ at all, especially for the land sites. The SLAB simulation ~~significantly improves EMD at two sites , (~~improves the EMD at four sites (SS > ~~5 %, relative to the base) ,~~ 0.1) while the MO simulation ~~significantly~~ improves the EMD at ~~four~~ three of the eight sites.

~~Based on these results, the MO simulation , which uses the NOAH LSM, MYNN PBL scheme and MO surface layer scheme (MO) was selected as the configuration for the NEWA production run~~Two other metrics are presented in Fig. 10: the RMSE for the wind speed (Fig. 10e) and the circular EMD (CEMD) for the wind direction (Fig. 10g) and their respective SS from the BASE simulation (Figs. 10f and 10h). As with the bias and EMD, the RMSE is primarily site dependent, with small differences between the simulations. The RMSE is lowest at Cabauw (1.6–2.1 $\mathrm{m\,s^{-1}}$) and highest at FINO2 (2.4–2.7 $\mathrm{m\,s^{-1}}$). No simulation significantly improves the RMSE, but it remains unchanged at all sites in the MO, SLAB, NOAHMP-WRF and the all the varied roughness ensembles. The CEMD of the wind direction is variable among sites and simulations, but all the values are small (1.9–7.1°). For this metric, the MO and SLAB show similar values than the BASE simulation.

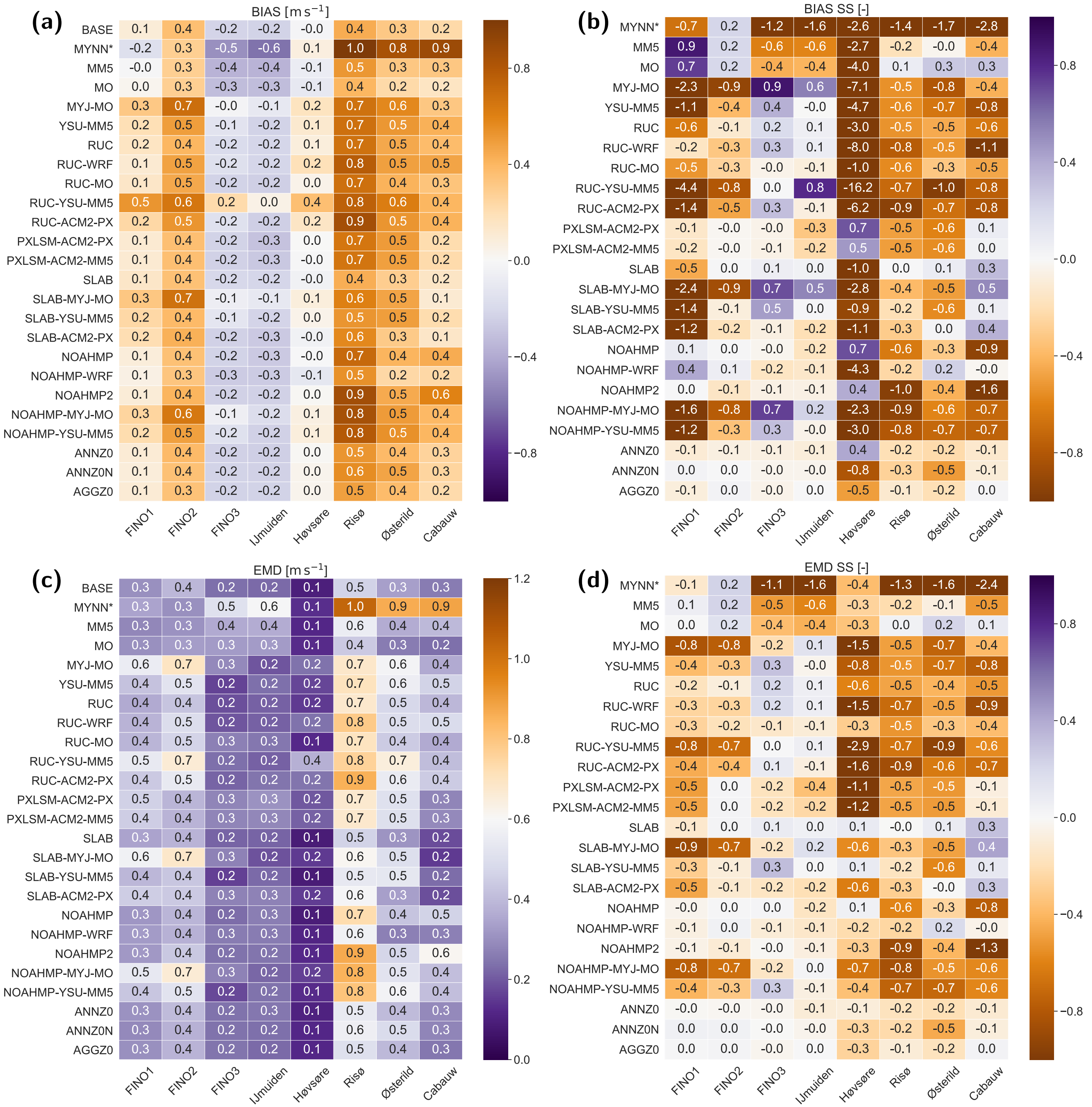


**Figure 10.** Evaluation metrics: (a) bias [$\mathrm{m\,s^{-1}}$], (b) bias SS [-], (c) EMD [$\mathrm{m\,s^{-1}}$], (d) EMD SS [-], (e) RMSE [$\mathrm{m\,s^{-1}}$], (f) RMSE SS [-], (g) CEMD [°], and (h) CEMD SS [-] between the observed and simulated wind speed and wind direction at the eight sites and underlined heights in Table 1 and the various sensitivity studies in the LSM/SL/PBL ensemble (Table 3). All SS are relative to the BASE simulation.

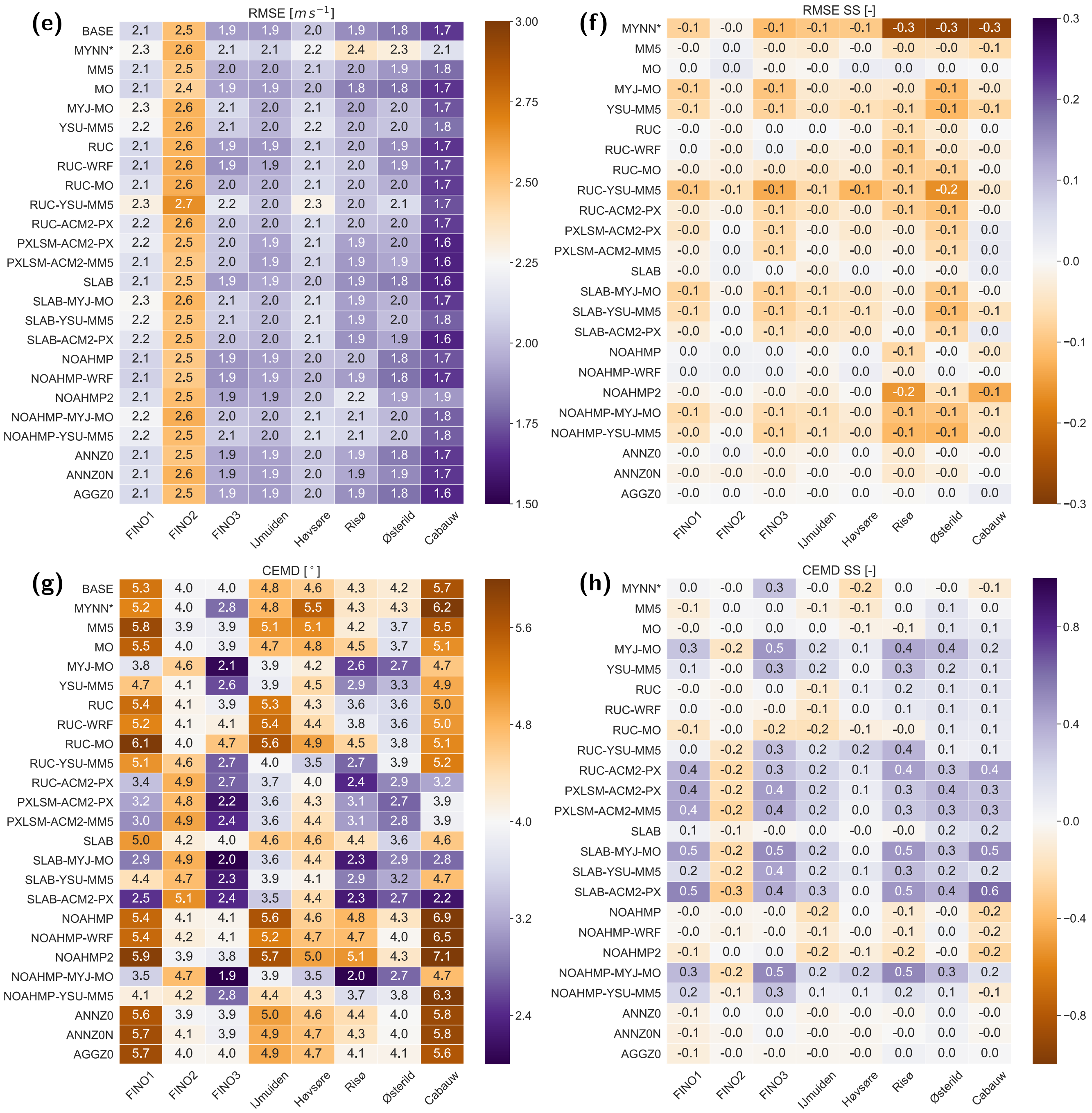


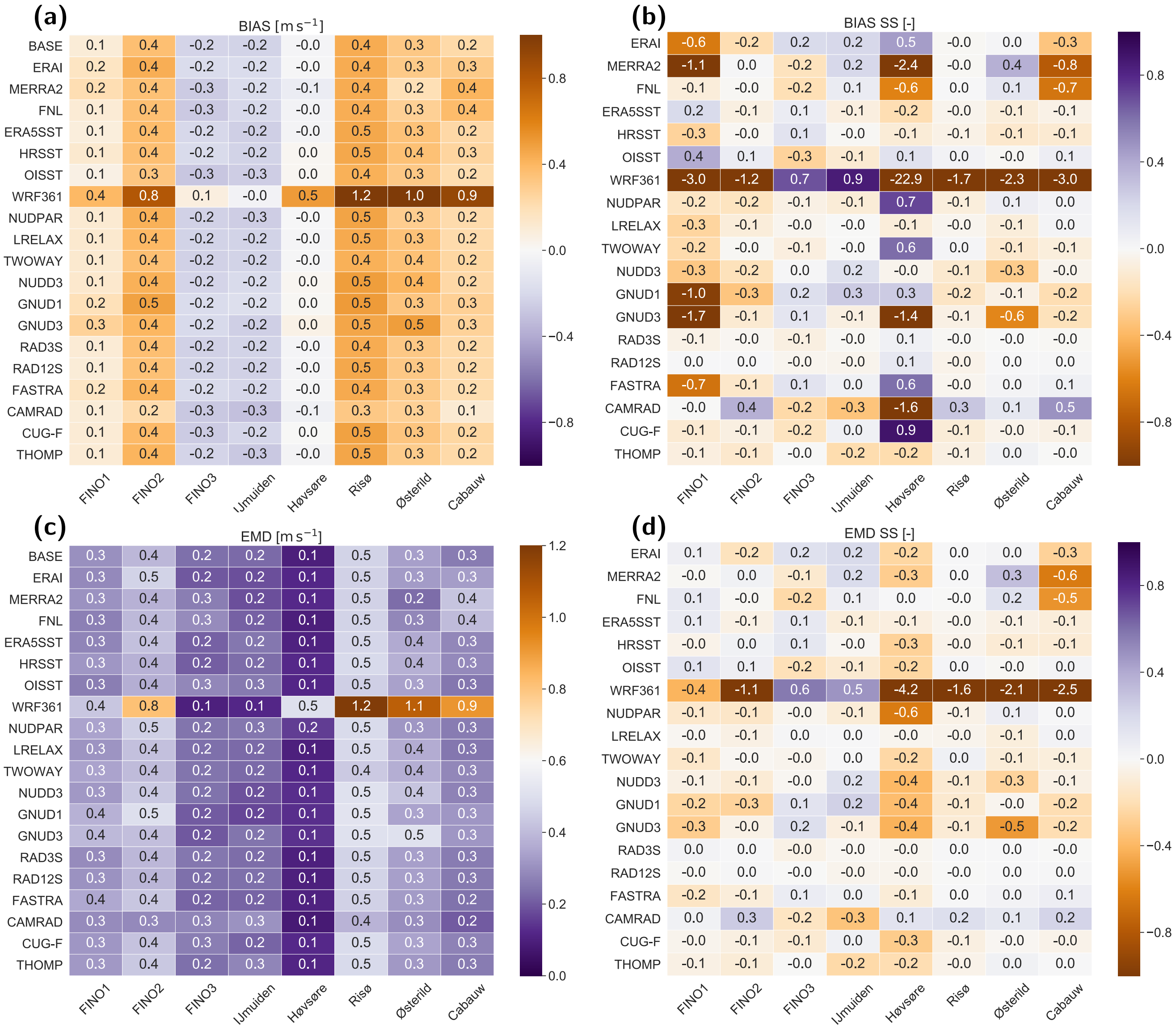
**Figure 10.** (Continued)

**Table 5.** Overview of other sensitivity experiments carried out. The meaning of the simulation code is explained in Table A2 in the appendix.

| run name | simulation code | changes to BASE run |
|---|---|---|
| Initial, boundary conditions and SST | | |
| ERAI | IIS81_2551040004 | ERA-Interim (Dee et al., 2011) forcing |
| MERRA2 | MMS81_2551040004 | MERRA2 (Gelaro et al., 2017) forcing |
| FNL | FFS81_2551040004 | FNL (NCAR, 2000) forcing |
| ERA5SST | EEE81_2551040004 | ERA5 SST |
| HRSST | EEH81_2551040004 | NOAA HRSST (Gemmill et al., 2007) |
| OISST | EEO81_2551040004 | NOAA RTGSST ~~Reynolds et al. (2010)~~ (Reynolds et al., 2010) |
| Model dynamics | | |
| WRFV361 | EES61_2550040004 | WRF version 3.6.1 |
| NUDPAR | EES81_2551040004_A | lower wavelength in nudging |
| LRELAX | EES81_2551040004_B | larger relaxation zone |
| TWOWAY | EES81_2551040004_C | two-way nesting |
| NUDD3 | EES81_2551040004_J | spectral nudging D1, D2, D3 |
| GNUD1 | EES81_2551040004_I | grid nudging D1 |
| GNUD3 | EES81_2551040004_H | grid nudging D1, D2, D3 |
| Other physics | | |
| RAD3S | EES81_2551040004_D | radiation $\Delta t = 3$ s in all domains |
| RAD12S | EES81_2551040004_E | radiation $\Delta t = 12$ s in all domains |
| FASTRA | EES81_2551040024 | fast RRTMG code |
| CAMRAD | EES81_2551040003 | CAM radiation (Collins et al., 2004) |
| CUG-F | EES81_2551040304 | Grell-Freitas (Grell and Freitas, 2014) CU scheme |
| THOMP | EES81_2551080004 | Thompson cloud physics (Thompson et al., 2012) + icing |

## 5.5 Other sensitivity experiments

A second set of sensitivity experiments was carried out to identify other factors that could potentially be important for the simulation of wind speed and direction within the WRF model. These experiments are listed in Table 5 and can be grouped into three main categories. First, we tested the impact of various initial and boundary conditions, by using ERA-Interim, MERRA2 and FNL fields as forcing. The effect of various sources of sea surface temperature (SST) was also tested. In a second set, we tested other model dynamics including the effect of spectral versus grid nudging, enlarging the lateral boundary zone, changing the wavelength of the minimum spectral nudging length and enabling 2-way nesting. The third set of experiments tested other model physics not related to the surface and PBL, that is radiation, cumulus convection and explicit moisture schemes.



(a) BIAS [m s$^{-1}$]

| | FINO1 | FINO2 | FINO3 | IJmuiden | Høvsøre | Risø | Østerild | Cabauw |
|---|---|---|---|---|---|---|---|---|
| BASE | 0.1 | 0.4 | -0.2 | -0.2 | -0.0 | 0.4 | 0.3 | 0.2 |
| ERAI | 0.2 | 0.4 | -0.2 | -0.2 | -0.0 | 0.4 | 0.3 | 0.3 |
| MERRA2 | 0.2 | 0.4 | -0.3 | -0.2 | -0.1 | 0.4 | 0.2 | 0.4 |
| FNL | 0.1 | 0.4 | -0.3 | -0.2 | -0.0 | 0.4 | 0.3 | 0.4 |
| ERA5SST | 0.1 | 0.4 | -0.2 | -0.2 | -0.0 | 0.5 | 0.3 | 0.2 |
| HRSST | 0.1 | 0.4 | -0.2 | -0.2 | 0.0 | 0.5 | 0.4 | 0.3 |
| OISST | 0.1 | 0.3 | -0.3 | -0.3 | 0.0 | 0.4 | 0.3 | 0.2 |
| WRF361 | 0.4 | 0.8 | 0.1 | -0.0 | 0.5 | 1.2 | 1.0 | 0.9 |
| NUDPAR | 0.1 | 0.4 | -0.2 | -0.3 | -0.0 | 0.5 | 0.3 | 0.2 |
| LRELAX | 0.1 | 0.4 | -0.2 | -0.2 | -0.0 | 0.5 | 0.3 | 0.2 |
| TWOWAY | 0.1 | 0.4 | -0.2 | -0.2 | -0.0 | 0.4 | 0.4 | 0.2 |
| NUDD3 | 0.1 | 0.4 | -0.2 | -0.2 | 0.0 | 0.5 | 0.4 | 0.2 |
| GNUD1 | 0.2 | 0.5 | -0.2 | -0.2 | 0.0 | 0.5 | 0.3 | 0.3 |
| GNUD3 | 0.3 | 0.4 | -0.2 | -0.2 | 0.0 | 0.5 | 0.5 | 0.3 |
| RAD3S | 0.1 | 0.4 | -0.2 | -0.2 | -0.0 | 0.4 | 0.3 | 0.2 |
| RAD12S | 0.1 | 0.4 | -0.2 | -0.2 | -0.0 | 0.5 | 0.3 | 0.2 |
| FASTRA | 0.2 | 0.4 | -0.2 | -0.2 | -0.0 | 0.4 | 0.3 | 0.2 |
| CAMRAD | 0.1 | 0.2 | -0.3 | -0.3 | -0.1 | 0.3 | 0.3 | 0.1 |
| CUG-F | 0.1 | 0.4 | -0.3 | -0.2 | 0.0 | 0.5 | 0.3 | 0.2 |
| THOMP | 0.1 | 0.4 | -0.2 | -0.3 | -0.0 | 0.5 | 0.3 | 0.2 |

(b) BIAS SS [-]

| | FINO1 | FINO2 | FINO3 | IJmuiden | Høvsøre | Risø | Østerild | Cabauw |
|---|---|---|---|---|---|---|---|---|
| ERAI | -0.6 | -0.2 | 0.2 | 0.2 | 0.5 | -0.0 | 0.0 | -0.3 |
| MERRA2 | -1.1 | 0.0 | -0.2 | 0.2 | -2.4 | -0.0 | 0.4 | -0.8 |
| FNL | -0.1 | -0.0 | -0.2 | 0.1 | -0.6 | 0.0 | 0.1 | -0.7 |
| ERA5SST | 0.2 | -0.1 | 0.1 | -0.1 | -0.2 | -0.0 | -0.1 | -0.1 |
| HRSST | -0.3 | -0.0 | 0.1 | -0.0 | -0.1 | -0.1 | -0.1 | -0.1 |
| OISST | 0.4 | 0.1 | -0.3 | -0.1 | 0.1 | 0.0 | -0.0 | 0.1 |
| WRF361 | -3.0 | -1.2 | 0.7 | 0.9 | -22.9 | -1.7 | -2.3 | -3.0 |
| NUDPAR | -0.2 | -0.2 | -0.1 | -0.1 | 0.7 | -0.1 | 0.1 | 0.0 |
| LRELAX | -0.3 | -0.1 | -0.0 | -0.0 | -0.1 | -0.0 | -0.1 | 0.0 |
| TWOWAY | -0.2 | -0.0 | -0.1 | -0.0 | 0.6 | 0.0 | -0.1 | -0.1 |
| NUDD3 | -0.3 | -0.2 | 0.0 | 0.2 | -0.0 | -0.1 | -0.3 | -0.0 |
| GNUD1 | -1.0 | -0.3 | 0.2 | 0.3 | 0.3 | -0.2 | -0.1 | -0.2 |
| GNUD3 | -1.7 | -0.1 | 0.1 | -0.1 | -1.4 | -0.1 | -0.6 | -0.2 |
| RAD3S | -0.1 | -0.0 | -0.1 | -0.0 | 0.1 | -0.0 | -0.0 | -0.0 |
| RAD12S | 0.0 | 0.0 | -0.0 | -0.0 | 0.1 | -0.0 | 0.0 | 0.0 |
| FASTRA | -0.7 | -0.1 | 0.1 | 0.0 | 0.6 | -0.0 | 0.0 | 0.1 |
| CAMRAD | -0.0 | 0.4 | -0.2 | -0.3 | -1.6 | 0.3 | 0.1 | 0.5 |
| CUG-F | -0.1 | -0.1 | -0.2 | 0.0 | 0.9 | -0.1 | -0.0 | -0.1 |
| THOMP | -0.1 | -0.1 | -0.0 | -0.2 | -0.2 | -0.1 | 0.0 | -0.0 |

(c) EMD [m s$^{-1}$]

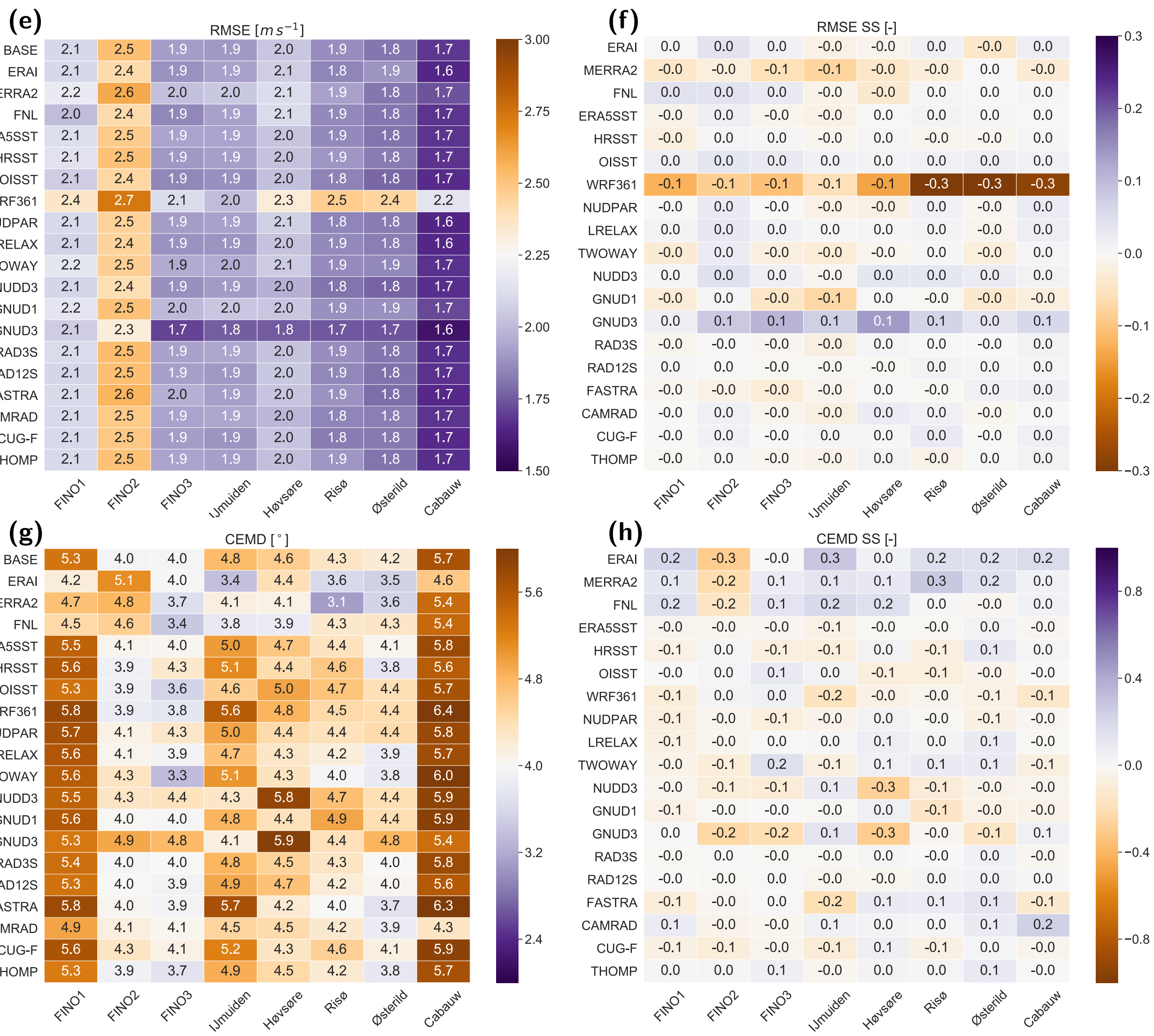
| | FINO1 | FINO2 | FINO3 | IJmuiden | Høvsøre | Risø | Østerild | Cabauw |
|---|---|---|---|---|---|---|---|---|
| BASE | 0.3 | 0.4 | 0.2 | 0.2 | 0.1 | 0.5 | 0.3 | 0.3 |
| ERAI | 0.3 | 0.5 | 0.2 | 0.2 | 0.1 | 0.5 | 0.3 | 0.3 |
| MERRA2 | 0.3 | 0.4 | 0.3 | 0.2 | 0.1 | 0.5 | 0.2 | 0.4 |
| FNL | 0.3 | 0.4 | 0.3 | 0.2 | 0.1 | 0.5 | 0.3 | 0.4 |
| ERA5SST | 0.3 | 0.4 | 0.2 | 0.2 | 0.1 | 0.5 | 0.4 | 0.3 |
| HRSST | 0.3 | 0.4 | 0.2 | 0.2 | 0.1 | 0.5 | 0.4 | 0.3 |
| OISST | 0.3 | 0.4 | 0.3 | 0.3 | 0.1 | 0.5 | 0.3 | 0.3 |
| WRF361 | 0.4 | 0.8 | 0.1 | 0.1 | 0.5 | 1.2 | 1.1 | 0.9 |
| NUDPAR | 0.3 | 0.5 | 0.2 | 0.3 | 0.2 | 0.5 | 0.3 | 0.3 |
| LRELAX | 0.3 | 0.4 | 0.2 | 0.2 | 0.1 | 0.5 | 0.4 | 0.3 |
| TWOWAY | 0.3 | 0.4 | 0.2 | 0.2 | 0.1 | 0.4 | 0.4 | 0.3 |
| NUDD3 | 0.3 | 0.4 | 0.2 | 0.2 | 0.1 | 0.5 | 0.4 | 0.3 |
| GNUD1 | 0.4 | 0.5 | 0.2 | 0.2 | 0.1 | 0.5 | 0.3 | 0.3 |
| GNUD3 | 0.4 | 0.4 | 0.2 | 0.2 | 0.1 | 0.5 | 0.5 | 0.3 |
| RAD3S | 0.3 | 0.4 | 0.2 | 0.2 | 0.1 | 0.5 | 0.3 | 0.3 |
| RAD12S | 0.3 | 0.4 | 0.2 | 0.2 | 0.1 | 0.5 | 0.3 | 0.3 |
| FASTRA | 0.4 | 0.4 | 0.2 | 0.2 | 0.1 | 0.5 | 0.3 | 0.2 |
| CAMRAD | 0.3 | 0.3 | 0.3 | 0.3 | 0.1 | 0.4 | 0.3 | 0.2 |
| CUG-F | 0.3 | 0.4 | 0.3 | 0.2 | 0.1 | 0.5 | 0.3 | 0.3 |
| THOMP | 0.3 | 0.4 | 0.2 | 0.3 | 0.1 | 0.5 | 0.3 | 0.3 |

(d) EMD SS [-]

| | FINO1 | FINO2 | FINO3 | IJmuiden | Høvsøre | Risø | Østerild | Cabauw |
|---|---|---|---|---|---|---|---|---|
| ERAI | 0.1 | -0.2 | 0.2 | 0.2 | -0.2 | 0.0 | 0.0 | -0.3 |
| MERRA2 | -0.0 | 0.0 | -0.1 | 0.2 | -0.3 | 0.0 | 0.3 | -0.6 |
| FNL | 0.1 | -0.0 | -0.2 | 0.1 | 0.0 | -0.0 | 0.2 | -0.5 |
| ERA5SST | 0.1 | -0.1 | 0.1 | -0.1 | -0.1 | -0.0 | -0.1 | -0.1 |
| HRSST | -0.0 | 0.0 | 0.1 | -0.0 | -0.3 | -0.0 | -0.1 | -0.1 |
| OISST | 0.1 | 0.1 | -0.2 | -0.1 | -0.2 | 0.0 | -0.0 | 0.0 |
| WRF361 | -0.4 | -1.1 | 0.6 | 0.5 | -4.2 | -1.6 | -2.1 | -2.5 |
| NUDPAR | -0.1 | -0.1 | -0.0 | -0.1 | -0.6 | -0.1 | 0.1 | 0.0 |
| LRELAX | -0.0 | -0.1 | 0.0 | 0.0 | -0.0 | -0.0 | -0.1 | 0.0 |
| TWOWAY | -0.1 | -0.0 | -0.0 | 0.0 | -0.2 | 0.0 | -0.1 | -0.1 |
| NUDD3 | -0.1 | -0.1 | -0.0 | 0.2 | -0.4 | -0.1 | -0.3 | -0.1 |
| GNUD1 | -0.2 | -0.3 | 0.1 | 0.2 | -0.4 | -0.1 | -0.0 | -0.2 |
| GNUD3 | -0.3 | -0.0 | 0.2 | -0.1 | -0.4 | -0.1 | -0.5 | -0.2 |
| RAD3S | 0.0 | 0.0 | -0.0 | -0.0 | -0.0 | 0.0 | 0.0 | -0.0 |
| RAD12S | -0.0 | 0.0 | -0.0 | -0.0 | -0.0 | -0.0 | 0.0 | 0.0 |
| FASTRA | -0.2 | -0.1 | 0.1 | 0.0 | -0.1 | 0.0 | 0.0 | 0.1 |
| CAMRAD | 0.0 | 0.3 | -0.2 | -0.3 | 0.1 | 0.2 | 0.1 | 0.2 |
| CUG-F | -0.0 | -0.1 | -0.1 | 0.0 | -0.3 | -0.1 | -0.0 | -0.0 |
| THOMP | -0.1 | -0.1 | -0.0 | -0.2 | -0.2 | -0.0 | 0.0 | 0.0 |

**Figure 11.** Evaluation metrics: (a) bias [m s$^{-1}$], (b) bias SS [-], (c) EMD [m s$^{-1}$], (d) EMD SS [-], (e) RMSE [m s$^{-1}$], (f) RMSE SS [-], (g) CEMD [°], and (h) CEMD SS [-] between the observed and simulated wind speed at the eight sites and underlined heights in Table 1 and the various sensitivity studies in the other ensemble (Table 5). All SS are relative to the BASE simulation.

Figure ~~?? shows the bias and bias differences compared to the BASE~~ 11 shows all the evaluation metrics of the wind speed and wind direction for each simulation for the ~~additional~~ other set of sensitivity experiments. In contrast to the LSM/PBL/SL ensemble members, for the bias (Figure 11a) these simulations provide results that are very similar to the BASE simulation, except for the ~~CAMRAD simulation, which was run replacing the usual RRTM radiation scheme (Mlawer et al., 1997) with the~~



(e) RMSE [$m\,s^{-1}$]

| | FINO1 | FINO2 | FINO3 | IJmuiden | Høvsøre | Risø | Østerild | Cabauw |
|---|---|---|---|---|---|---|---|---|
| BASE | 2.1 | 2.5 | 1.9 | 1.9 | 2.0 | 1.9 | 1.8 | 1.7 |
| ERAI | 2.1 | 2.4 | 1.9 | 1.9 | 2.1 | 1.8 | 1.9 | 1.6 |
| MERRA2 | 2.2 | 2.6 | 2.0 | 2.0 | 2.1 | 1.9 | 1.8 | 1.7 |
| FNL | 2.0 | 2.4 | 1.9 | 1.9 | 2.1 | 1.9 | 1.8 | 1.7 |
| ERA5SST | 2.1 | 2.5 | 1.9 | 1.9 | 2.0 | 1.9 | 1.8 | 1.7 |
| HRSST | 2.1 | 2.5 | 1.9 | 1.9 | 2.0 | 1.9 | 1.8 | 1.7 |
| OISST | 2.1 | 2.4 | 1.9 | 1.9 | 2.0 | 1.8 | 1.8 | 1.7 |
| WRF361 | 2.4 | 2.7 | 2.1 | 2.0 | 2.3 | 2.5 | 2.4 | 2.2 |
| NUDPAR | 2.1 | 2.5 | 1.9 | 1.9 | 2.1 | 1.8 | 1.8 | 1.6 |
| LRELAX | 2.1 | 2.4 | 1.9 | 1.9 | 2.0 | 1.9 | 1.8 | 1.6 |
| TWOWAY | 2.2 | 2.5 | 1.9 | 2.0 | 2.1 | 1.9 | 1.9 | 1.7 |
| NUDD3 | 2.1 | 2.4 | 1.9 | 1.9 | 2.0 | 1.8 | 1.8 | 1.7 |
| GNUD1 | 2.2 | 2.5 | 2.0 | 2.0 | 2.0 | 1.9 | 1.9 | 1.7 |
| GNUD3 | 2.1 | 2.3 | 1.7 | 1.8 | 1.8 | 1.7 | 1.7 | 1.6 |
| RAD3S | 2.1 | 2.5 | 1.9 | 1.9 | 2.0 | 1.9 | 1.8 | 1.7 |
| RAD12S | 2.1 | 2.5 | 1.9 | 1.9 | 2.0 | 1.9 | 1.8 | 1.7 |
| FASTRA | 2.1 | 2.6 | 2.0 | 1.9 | 2.0 | 1.9 | 1.8 | 1.7 |
| CAMRAD | 2.1 | 2.5 | 1.9 | 1.9 | 2.0 | 1.8 | 1.8 | 1.7 |
| CUG-F | 2.1 | 2.5 | 1.9 | 1.9 | 2.0 | 1.8 | 1.8 | 1.7 |
| THOMP | 2.1 | 2.5 | 1.9 | 1.9 | 2.0 | 1.9 | 1.8 | 1.7 |

(f) RMSE SS [-]

| | FINO1 | FINO2 | FINO3 | IJmuiden | Høvsøre | Risø | Østerild | Cabauw |
|---|---|---|---|---|---|---|---|---|
| ERAI | 0.0 | 0.0 | 0.0 | -0.0 | -0.0 | 0.0 | -0.0 | 0.0 |
| MERRA2 | -0.0 | -0.0 | -0.1 | -0.1 | -0.0 | -0.0 | 0.0 | -0.0 |
| FNL | 0.0 | 0.0 | 0.0 | -0.0 | -0.0 | 0.0 | 0.0 | 0.0 |
| ERA5SST | -0.0 | 0.0 | -0.0 | -0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| HRSST | -0.0 | 0.0 | 0.0 | -0.0 | 0.0 | -0.0 | -0.0 | 0.0 |
| OISST | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| WRF361 | -0.1 | -0.1 | -0.1 | -0.1 | -0.1 | -0.3 | -0.3 | -0.3 |
| NUDPAR | -0.0 | 0.0 | -0.0 | -0.0 | -0.0 | 0.0 | -0.0 | 0.0 |
| LRELAX | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | -0.0 | 0.0 |
| TWOWAY | -0.0 | 0.0 | -0.0 | -0.0 | -0.0 | 0.0 | -0.0 | 0.0 |
| NUDD3 | 0.0 | 0.0 | 0.0 | -0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| GNUD1 | -0.0 | 0.0 | -0.0 | -0.1 | 0.0 | -0.0 | -0.0 | -0.0 |
| GNUD3 | 0.0 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.0 | 0.1 |
| RAD3S | -0.0 | -0.0 | -0.0 | -0.0 | 0.0 | 0.0 | -0.0 | 0.0 |
| RAD12S | 0.0 | 0.0 | -0.0 | -0.0 | -0.0 | 0.0 | 0.0 | 0.0 |
| FASTRA | -0.0 | -0.0 | -0.0 | -0.0 | 0.0 | -0.0 | 0.0 | 0.0 |
| CAMRAD | -0.0 | 0.0 | -0.0 | -0.0 | 0.0 | 0.0 | -0.0 | 0.0 |
| CUG-F | -0.0 | 0.0 | -0.0 | 0.0 | 0.0 | 0.0 | -0.0 | 0.0 |
| THOMP | -0.0 | 0.0 | -0.0 | -0.0 | 0.0 | -0.0 | 0.0 | 0.0 |

(g) CEMD [$^\circ$]

| | FINO1 | FINO2 | FINO3 | IJmuiden | Høvsøre | Risø | Østerild | Cabauw |
|---|---|---|---|---|---|---|---|---|
| BASE | 5.3 | 4.0 | 4.0 | 4.8 | 4.6 | 4.3 | 4.2 | 5.7 |
| ERAI | 4.2 | 5.1 | 4.0 | 3.4 | 4.4 | 3.6 | 3.5 | 4.6 |
| MERRA2 | 4.7 | 4.8 | 3.7 | 4.1 | 4.1 | 3.1 | 3.6 | 5.4 |
| FNL | 4.5 | 4.6 | 3.4 | 3.8 | 3.9 | 4.3 | 4.3 | 5.4 |
| ERA5SST | 5.5 | 4.1 | 4.0 | 5.0 | 4.7 | 4.4 | 4.1 | 5.8 |
| HRSST | 5.6 | 3.9 | 4.3 | 5.1 | 4.4 | 4.6 | 3.8 | 5.6 |
| OISST | 5.3 | 3.9 | 3.6 | 4.6 | 5.0 | 4.7 | 4.4 | 5.7 |
| WRF361 | 5.8 | 3.9 | 3.8 | 5.6 | 4.8 | 4.5 | 4.4 | 6.4 |
| NUDPAR | 5.7 | 4.1 | 4.3 | 5.0 | 4.4 | 4.4 | 4.4 | 5.8 |
| LRELAX | 5.6 | 4.1 | 3.9 | 4.7 | 4.3 | 4.2 | 3.9 | 5.7 |
| TWOWAY | 5.6 | 4.3 | 3.3 | 5.1 | 4.3 | 4.0 | 3.8 | 6.0 |
| NUDD3 | 5.5 | 4.3 | 4.4 | 4.3 | 5.8 | 4.7 | 4.4 | 5.9 |
| GNUD1 | 5.6 | 4.0 | 4.0 | 4.8 | 4.4 | 4.9 | 4.4 | 5.9 |
| GNUD3 | 5.3 | 4.9 | 4.8 | 4.1 | 5.9 | 4.4 | 4.8 | 5.4 |
| RAD3S | 5.4 | 4.0 | 4.0 | 4.8 | 4.5 | 4.3 | 4.0 | 5.8 |
| RAD12S | 5.3 | 4.0 | 3.9 | 4.9 | 4.7 | 4.2 | 4.0 | 5.6 |
| FASTRA | 5.8 | 4.0 | 3.9 | 5.7 | 4.2 | 4.0 | 3.7 | 6.3 |
| CAMRAD | 4.9 | 4.1 | 4.1 | 4.5 | 4.5 | 4.2 | 3.9 | 4.3 |
| CUG-F | 5.6 | 4.3 | 4.1 | 5.2 | 4.3 | 4.6 | 4.1 | 5.9 |
| THOMP | 5.3 | 3.9 | 3.7 | 4.9 | 4.5 | 4.2 | 3.8 | 5.7 |

(h) CEMD SS [-]

| | FINO1 | FINO2 | FINO3 | IJmuiden | Høvsøre | Risø | Østerild | Cabauw |
|---|---|---|---|---|---|---|---|---|
| ERAI | 0.2 | -0.3 | -0.0 | 0.3 | 0.0 | 0.2 | 0.2 | 0.2 |
| MERRA2 | 0.1 | -0.2 | 0.1 | 0.1 | 0.1 | 0.3 | 0.2 | 0.0 |
| FNL | 0.2 | -0.2 | 0.1 | 0.2 | 0.2 | 0.0 | -0.0 | 0.0 |
| ERA5SST | -0.0 | -0.0 | -0.0 | -0.1 | -0.0 | -0.0 | 0.0 | -0.0 |
| HRSST | -0.1 | 0.0 | -0.1 | -0.1 | 0.0 | -0.1 | 0.1 | 0.0 |
| OISST | -0.0 | 0.0 | 0.1 | 0.0 | -0.1 | -0.1 | -0.0 | -0.0 |
| WRF361 | -0.1 | 0.0 | 0.0 | -0.2 | -0.0 | -0.0 | -0.1 | -0.1 |
| NUDPAR | -0.1 | -0.0 | -0.1 | -0.0 | 0.0 | -0.0 | -0.1 | -0.0 |
| LRELAX | -0.1 | -0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.1 | -0.0 |
| TWOWAY | -0.0 | -0.1 | 0.2 | -0.1 | 0.1 | 0.1 | 0.1 | -0.1 |
| NUDD3 | -0.0 | -0.1 | -0.1 | 0.1 | -0.3 | -0.1 | -0.0 | -0.0 |
| GNUD1 | -0.1 | -0.0 | -0.0 | -0.0 | 0.0 | -0.1 | -0.0 | -0.0 |
| GNUD3 | 0.0 | -0.2 | -0.2 | 0.1 | -0.3 | -0.0 | -0.1 | 0.1 |
| RAD3S | -0.0 | 0.0 | -0.0 | -0.0 | 0.0 | -0.0 | 0.0 | -0.0 |
| RAD12S | -0.0 | -0.0 | 0.0 | -0.0 | -0.0 | 0.0 | 0.0 | 0.0 |
| FASTRA | -0.1 | -0.0 | 0.0 | -0.2 | 0.1 | 0.1 | 0.1 | -0.1 |
| CAMRAD | 0.1 | -0.0 | -0.0 | 0.1 | 0.0 | 0.0 | 0.1 | 0.2 |
| CUG-F | -0.1 | -0.1 | -0.0 | -0.1 | 0.1 | -0.1 | 0.0 | -0.0 |
| THOMP | 0.0 | 0.0 | 0.1 | -0.0 | 0.0 | 0.0 | 0.1 | -0.0 |

**Figure 11.** (Continued)

~~CAM parameterisation scheme (Collins et al., 2004)~~WRF361, GNUD3 and CAMRAD simulations. Switching the source of initial and boundary conditions to ERA-Interim, MERRA2 or FNL has very small impact to the bias~~ratios. The only significant change is at Cabauw, where the biases are increased by 1–2 % by using any of these three other forcing data.~~. Changing the source of SST has ~~an insignificant~~ negligible effect for all of the offshore masts. Most of the changes to the dynamic settings have very small consequences to the bias. The only ~~significant~~ larger change is the use of grid nudging in all three WRF domains, simulation GNUD3. The biases increase ~~in six~~ (SS $\geq$ 0.1) in seven of the eight sites ~~by 0.2–2.0 %~~ when this setting is activated, probably because of the slow down of the winds in the ERA5 reanalysis over land (see Fig. 9 of Dörenkämper et al., 2020). Interestingly, using this setting ~~significantly increases the correlation~~ decreases the RMSE between the simulated and observed time series (~~not shown), but at the expense of increased biases.~~Fig. 11f). Since the intent of the NEWA atlas is to provide an accurate description of the wind climatology, distribution metrics (e.g. EMD) are considered more important than time-dependent ones (e.g. RMSE).

~~(a) Biases %and (b) relative change in the biases from BASE simulation |BIAS| − |BIAS_B|, %between the observed and simulated wind speed at the eight sites and the various sensitivity studies in the non-PBL ensemble (Table 5).~~

~~(a) EMD relative to the observed wind speed %and (b) relative change in the EMD from the BASE simulation (EMD − EMD_B)/EMD_ %between the observed and simulated wind speed at the eight sites and the various sensitivity studies in the non-PBL ensemble (Table 5).~~

~~Similarly, the relative EMD~~ The EMD and EMD SS for this set of sensitivity experiments was calculated and the results are shown in Figure ~~**??**. The left panel shows the EMD relative to the observed wind speed and right panel the relative improvement of this metric with respect to the baseline~~11c and 11d. The conclusions about the usefulness of ~~EMD metric for FINO1 and Høvsøre~~ the EMD apply here as well. For instance, at FINO1 the ERAI and MERRA2 runs show an increase in bias compared to the BASE, while the EMD values show that these runs actually have more similar wind speed distribution to observations than the BASE. Otherwise, the EMD metric confirms the conclusions described earlier about the small impact of all of these changes and the relative decrease in quality when using grid-nudging (GNUD3).

Other dynamical and physics options such as the radiation time step (RAD3S or RAD12S), the change of nudging constants (NUDPAR, LRELAX) or nesting (TWOWAY) show very small influence to nearly all the evaluation metrics. As with the LSM/PBL/SL, the simulations show very small differences in RMSE and CEMD with respect to the BASE, except for GNUD3 as mentioned above, which shows improved RMSE. It is also interesting to note that the ERAI, MERRA2 and FNL simulations show small improvements in the CEMD at many sites compared to the BASE simulation. The reason for this behaviour is not well understood.

In conclusion, many other changes to the WRF model settings have inconsequential effects to the simulation of the wind speed and direction at wind turbine hub height. The use of grid nudging in all domains (simulations GNUD3) improves the RMSE at most sites, but has a negative effect to the BIAS, EMD and CEMD. The change in radiation parameterisation has a small effect in relation to the BASE simulation. ~~Unfortunately, we did not run a simulation with the MO and CAMRAD together, so it is not possible to assess if that simulation would have been more accurate.~~ Because the effect is small and the CAM radiation parameterisation (Collins et al., 2004) is more expensive in terms of computational resources, it was ultimately ~~decided to keep the NOAH-MYNN-MO setup as the choice for the~~ not used in the NEWA production run.

### 5.6 ~~Domain~~ Sensitivity to domain size

An additional decision to be made regarding the NEWA mesoscale simulations was the domain configuration; that is using a single large domain or several small domains to cover Europe. From a pure computational perspective, one single domain is more efficient, because the WRF model code scales better with larger domains (Kruse et al., 2013) and there is only data

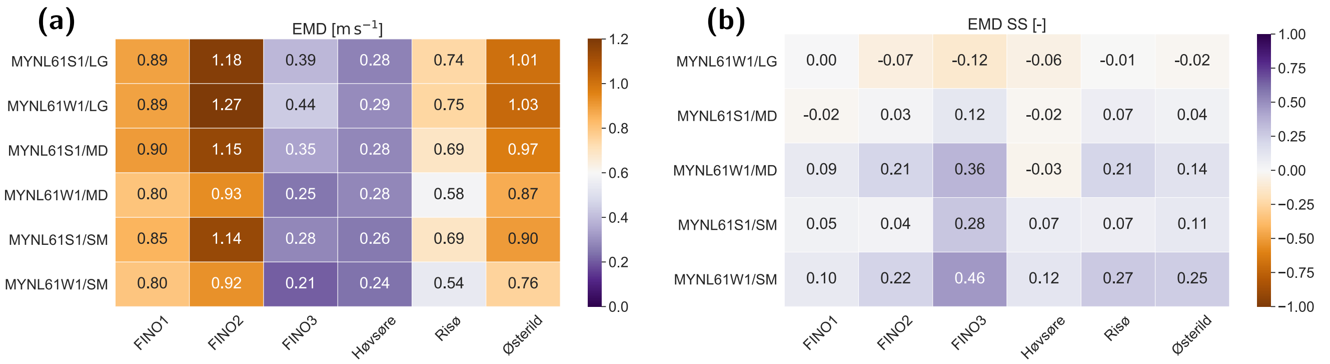


**Figure 12.** ~~Relative bias~~ (a) EMD [~~%~~m s$^{-1}$] ~~in annual mean~~ and (b) EMD SS [-] between the observed and simulated wind speed for ~~each~~ the six inner sites and underlined heights in Table 1 for the six domain size ~~and~~ experiments. The EMD SS is compared to that of the EMD of the MYNL61W1/SM simulation~~length for FINO1, FINO2, FINO3 (top) and Høvsøre, Østerild and Risø(bottom)~~.

from one domain to post-process. However, the output files are very large and the simulation needs to be completed before post-processing can begin. The limiting factor here is the scratch space available at modern HPC systems that is typically not more than 100 TB. Furthermore, large areas outside of the region of interest (the NEWA domain, see Section 1) would be simulated, that is parts of the Atlantic Ocean, the Norwegian Sea and non-EU countries in Eastern Europe, ~~thus a substantial amount of computational resources would be wasted~~which were outside of the planned New European Wind Atlas. Apart from these technical questions, it was unknown how the domain size influences the quality of the simulated fields.

To study the sensitivity of the simulated wind speed, we carried out simulations for three differently sized domains over the North Sea using the same setup and resolution as in Section 5.1, but for 2016, when data availability was larger. The number of grid points in the inner domain in these simulations are: small (SM) 121 × 121, medium (MD) 241 × 241 and large (LG) 481 × 481, which correspond to square domains with edge lengths on the WRF model projection equivalent to 360 km, 720 km, and 1440 km, respectively. The three domains are centred at the same coordinates and only differ by the number of grid points. The size of the boundary zone, in grid cells, between D1 and D2 and D2 and D3 is kept the same. Two sets of WRF model simulations were done for each of the three domains~~with daily and weekly runs, analogous to~~ , MYNL61S1 and MYNL61W1 in Table 2. For evaluating the results of the simulations we use the same data as in Section 5.1, but only six of the masts are contained within the SM domain.

Figure 12 shows the ~~biases~~ EMD between the various WRF model simulations and the observations for the ~~6 sites. The biases~~ six sites. It should be noted that these EMD values are larger than those previously shown for the LSM/SL/PBL and other ensemble. The reason for this is that the simulations were carried out with WRF V3.6.1 (see brown line in Fig. 11c) and the measurements at FINO1 and FINO2 are more affected by the presence of neighbouring wind farms in 2016 than what they were in 2015. The EMD from all simulations are summarised as follows: for five out of six sites the ~~MYNL61S1~~MYNL61W1/LG simulation have the largest ~~biases~~EMD, and for all sites the MYNL61W1/SM simulation has the smallest ~~biases~~EMD. Similar results (not shown) emerge for the ~~correlation and the RMSE. Particularly at Høvsøre, the bias decreases from 2.9 to 1.6 to 0.6 in the MYNL61W1 LG, MD, and SM simulations, respectively.~~

bias and RMSE. In conclusion, for this region, biases in mean wind speed are influenced by the size (and possibly also location~~, not shown~~) of the domain, smaller domains have generally lower wind speeds and thus lower biases. This effect is most pronounced in the week-long ~~and “nudged ” simulations. Time correlations decrease (and RMSE increase)~~ nudged simulations. The RMSE increases with increasing domain size and integration time.

The results from these experiments guided the design of the NEWA domains for the production run. Instead of a single, or a few, very large domains, we chose to conduct the simulations in a rather large number of medium-sized domains. While generating different time series, overlapping areas in simulations generally show similar wind climates (Witha et al., 2019, Section 2.1.3). We decided ~~, however,~~ against very small domains~~. In terms of accuracy they would probably perform better than our chosen configuration, but~~ , which would add to the needed computational resources and the extra effort of dealing with hundreds of model domains and because most countries would be covered by multiple domains ~~, which would face overlapping issues.~~ It was desired that each country should be covered by only one domain to avoid ~~these issues~~edge differences in overlapping domains. The final domain configuration, which is presented in this paper’s companion (Dörenkämper et al., 2020), fulfils this requirement for all countries except Norway, Sweden and Finland which are so elongated that a correspondingly large domain would be detrimental to the accuracy of the results.

### 5.7 Summary of the sensitivity experiments

A long list of sensitivity studies were carried out to identify ~~the ideal~~ a suitable configuration for the NEWA production run. Here is a summary of the findings:

1. In the initial sensitivity experiments, the largest differences in annual mean wind speed at 100 m ~~a.g.l. vs height in figures~~ in Fig. 4 and AGL/AMSL in Table 2 are between simulations using two PBL schemes (MYNN and YSU) and coincide with regions of high surface roughness in all domains over Europe. Over the sea, the differences could be traced to differences in atmospheric stability, but were modest.

2. The weekly simulations using spectral nudging tend to perform better (lower ~~biases and higher correlations~~bias, EMD and RMSE) for the eight sites in northern Europe. The simulation using the MYNN scheme in the WRF model version 3.6.1 outperformed the simulations using the YSU scheme in this region.

3. The use of the WRF model V3.8.1 and the MYNN scheme increased the biases compared to observations at nearly all sites and most levels. A couple of settings, mynn_mixlength=0 and COARE_OPT=3.0, turn the MYNN scheme nearly back to the conditions in the WRF model version 3.6.1. However, above 100 m a.g.l. the modified MYNN scheme gives lower wind speeds than the one in WRF V3.6.1 at all sites.

4. A series of 25 experiments varying the land, PBL and surface layer scheme (LSM/PBL/SL) shows a spread in the mean wind speed of about $1\,\mathrm{m\,s^{-1}}$~~over the ocean and 1.5 $\mathrm{m\,s^{-1}}$ over land~~. When comparing to wind speed observations,

most LSM/PBL/SL ensembles show negative biases over the ocean (except for FINO2) and positive over land, which are more consistent between sites than LSM/PBL/SL combinations. This likely reflects misrepresentation of the land surface around each site than deficiencies in the LSM/PBL/SL schemes themselves.

5. Changes to the WRF model lookup table for surface roughness length have large consequences for the simulated wind speed, but is nearly invisible to the evaluation against the tall masts because these are located in areas away from those impacted by the changes.

6. The use of the EMD metric and the skill score helps clarify the comparison of the improvements between the various LSM/PBL/SL and the ~~baseline~~ BASE simulation, especially if the bias is small. No simulation improves the EMD for all sites, and very few simulations improve the EMD metric at all, especially for the land sites. However, the MO simulation, which uses the NOAH, MYNN and MO surface layer schemes, improves the results, SS $>$ ~~1 %~~ 0.1, at four of the eight sites and was finally chosen as the physical model configuration for the NEWA production run.

7. A set of additional sensitivity experiments ~~, change~~ changing the source of forcing data, SST, dynamic options and other physical parameterisations, shows smaller changes from the baseline simulation than the various LSM/PBL/SL experiments. Nearly all the changes have inconsequential effects to the simulation of the wind speed at wind turbine hub height. Only the simulation using the CAM radiation scheme showed improvements over the RRTMG scheme used in the baseline experiment. However, it was concluded that the modest improvements were not worth the additional expense of running this scheme in the production run.

8. A final set of experiments testing the effect of the size of the domain on the simulated wind speed error statistics showed that for a domain centred over Denmark, the simulations using the smaller domain have lower wind speeds which compare better to measurements and time correlations decrease with domain size. It is however unclear if this is a consequence of the domain size itself or the location of the main inflow boundary to the domain in the simulations.

## 6 Discussion and outlook

In the companion paper (Dörenkämper et al., 2020), we document the final model configuration and how we computed the final wind atlas, including a detailed description of the technical and practical aspects that went in to running the WRF simulations and the downscaling using the linearised microscale model WAsP (Troen and Petersen, 1989). This second paper also shows a comprehensive evaluation of each component of the NEWA model-chain using observations from a large set ($n = 291$) of tall masts located all over Europe. ~~We conclude~~ That work concludes that the NEWA wind ~~climates estimated by WRF and WAsP are~~ climate estimated by the WRF model simulations using the model configuration selected here is significantly more accurate than using the raw ERA5 ~~reanalysis data~~wind speed and direction which was used to force these simulations.

From the results of the sensitivity analysis, it is clear that some model setups have better scores for one validation metric (e.g. EMD) and some setups have better scores in other metrics (e.g. RMSE). Also, the performance of the setup varies considerably from station to station. In addition, one can clearly see that the performance at each station is systematic. At some sites all setups tend to perform better and at other sites worse. Thus, it was decided not to base configuration decisions on average metrics across the sites. From the results of the sensitivity analysis, several model setups could have been chosen (e.g. MO, SLAB or CAMRAD). The MO setup, which was ultimately chosen for the production runs (Dörenkämper et al., 2020), is one of the best performing, according to the verification results. It performs well both in terms of its wind speed distribution (EMD) and relatively well in a time-series sense (RMSE) compared to other model setups. It has the additional benefit of runs being numerically stable when compared to BASE, i.e., the runs failed less (this aspect was important, because of the necessity for a person to monitor and re-submit the runs), and MO also had a favourably small computational time when compared to other setups.

Many other details of the model setup have not been tested. For example, the sensitivity of the simulated wind climate to the number and location of the vertical levels in the PBL was not systematically tested. Previous studies (Hahmann et al., 2015) and earlier simulations in the NEWA project showed small impact, but these were conducted with a single PBL scheme. The dependence of the simulated wind climate on the grid spacing is also not explored here. Previous publications offshore (e.g. Floors et al., 2018b) and onshore (Rife and Davis, 2005; Gómez-Navarro et al., 2015; Smith et al., 2018) have done so, but the investigation on what is the ideal grid spacing of the mesoscale simulations when further downscaling is done also remains unknown. In a similar theme, the dependence of the WRF model simulation of the wind climate on the size and location of the computational domain also remains unresolved. Smaller domains in the WRF simulation tend to have smaller wind speed biases and higher RMSE compared to tall mast observations, but it was unclear if this was really a result of the size of the domain or rather the location of the boundaries in relation to the large-scale flow. More numerical experiments should be carried out to identify all these potential interactions.

As with any modelling study, some questions remain unresolved simply because of the expensive nature of the numerical experiments. For convenience and simplicity, we separated the sensitivity experiments dealing with LSM/PBL/SL and the other parameterisations changing only one scheme or parameter at a time. However, the experiment using the CAM radiation scheme had better verification statistics than the other simulations~~, but it~~. The use of the CAM scheme was not tested using the ~~final~~ preferred LSM/PBL/SL combination. ~~Therefore,~~ Thus, we suggest that a better way to go in this process ~~would be to sequentially~~ of model selection is to go through the changes and evaluation in a sequential way. But the number of ensemble members can rapidly become unmanageable. Algorithms in this direction are currently being applied for tuning Earth System Models (Li et al., 2019) and could perhaps be evaluated to best ~~optimal~~ determine the best suited WRF setups for different applications, not just wind resource assessment.

~~The dependence of the WRF model simulation of the wind climate on the size and location of the computational domain also remains unresolved. Smaller domains in the WRF simulation tend to have smaller wind speed biases and higher correlations compared to tall mast observations, but it was unclear if this was really a result of the size of the domain or rather the location of the boundaries in relation to the large-scale flow. More numerical experiments should be carried out to identify these potential interactions.~~

~~Finally, it would have been optimal to evaluate~~ Finally, we have not evaluated the results of the ensemble simulations with the large observational dataset used in the companion paper (Dörenkämper et al., 2020)~~and with the~~. We have also not evaluated the simulated wind climatologies when we use the full downscaling model chain. Nevertheless, the ~~results of the~~ evaluation of the production run with the data included there, support the performance of the configuration selected herein.

*Code availability.* The WRF model code is open source code and can be obtained from the WRF Model User's Page (http://www2.mmm.ucar.edu/wrf/users/, doi:10.5065/D6MK6B4K). For the NEWA production run we used WRF version 3.8.1. The code modifications as well as namelists, tables and domain files we used are available from the NEWA GitHub repository: https://github.com/newa-wind/Mesoscale and permanently indexed in Zenodo (Hahmann et al., 2020). The WRF namelists and tables for all the ensemble members are also available in the repositories. The code used in the calculation of EMD metric is available from: https://pypi.org/project/pyemd/

*Data availability.* The NEWA data is available from https://map.neweuropeanwindatlas.eu/. The forcing data for the mesoscale simulations are publicly available:
ERA5 - https://climate.copernicus.eu/climate-reanalysis,
OSTIA - http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=SST_GLO_SST_L4_NRT_OBSERVATIONS_010_001,
CORINE - https://land.copernicus.eu/pan-european/corine-land-cover,
ESA-CCI - http://cci.esa.int/data.
Some of the tall mast data used for the evaluation of the wind atlas is confidential and thus not publicly available.

**Table A1.** The WRF model setup common to all simulations and to the ~~baseline~~ BASE simulation.

| option | setting |
|---|---|
| **Common setup:** | |
| Model grid | D1/D2/D3 with 27 km / 9 km / 3 km horizontal grid spacing |
| | Lambert conformal grid projection |
| Terrain data | Global Multi-resolution Terrain Elevation Data 2010 at 30” (Danielson and Gesch, 2011) |
| Land use | CORINE land-cover classification (Copernicus Land Monitoring Service, 2019) |
| | ESA-CCI land-cover (Poulter et al., 2015) outside the CORINE domain |
| Vertical discretisation | 61 vertical levels with model top at 50 ~~HPa~~hPa. |
| Model levels | 20 model levels below 1 km |
| Diffusion | Simple diffusion (option 1), 2D deformation (option 4) |
| | 6th order positive definite numerical diffusion (option 2) |
| | No vertical damping |
| | Positive definite advection of moisture and scalars |
| **BASE setup:** | |
| Forcing data | ERA5 (Hersbach et al., 2020) reanalysis at 0.3° on pressure levels |
| Sea surface temperature | Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA, Donlon et al., 2012) |
| | fractional sea-ice activated |
| | Lake temperatures from time-averaged ERA5 ground temperatures |
| Cloud micro-physics | WRF Single-Moment 5-class scheme (Hong et al., 2004) |
| Cumulus convection | Kain-Fritsch Scheme (Kain, 2004); D1 and D2 |
| PBL scheme | MYNN level 2.5 (Nakanishi and Niino, 2009) |
| Surface layer scheme | MYNN (Nakanishi and Niino, 2009) with mods (see text) |
| Land surface model | Unified Noah Land Surface Model (Tewari et al., 2004) |
| Shortwave and longwave radiation | RRTMG (Iacono et al., 2008) at 12 minute interval |
| Nesting | one way nesting with smooth (option 2) |
| Nudging | spectral nudging U, V, T and q on D1 |
| | above level 20, no PBL nudging |
| Nudging constant | 0.0003 $s^{-1}$ |
| Nudging wavelength | 14 (x) and 10 (y) equivalent to $6 \times \Delta x$ of ERA-Interim reanalysis grid spacing |

**Table A2.** Explanation for the various digits of the sensitivity experiments that refers to the code available in GitHub.

| digit | option | convention |
|---|---|---|
| 1 | IC/BC data | E: ERA5, I: ERA-Interim, M: MERRA2, C: CFSR2, F: FNL |
| 2 | Land IC | same as digit 1 (E, I, M, C, F), G: GLDAS |
| 3 | SST data | S: OSTIA, H: HRSST, O: OISST, or same as digit 1 (E, I, M, C, F) |
| 4 | WRF version | 6: WRFV3.6.1, 8: WRFV3.8.1 |
| 5 | Roughness option | 1: constant, 2: annual cycle, 3: aggregated |
| 6 | Separator underscore | |
| 7 | Land Surface Model | Code as in WRF: Thermal diffusion=1, NOAH=2, RUC=3, CLM4=4, PLX=7 |
| 8 | PBL scheme | Code as in WRF: YSU=1, MYJ=2, QNSE=4, MYNN2=5, MYNN3=6, ACM2=7 |
| 9 | Surface layer | Code as in WRF: Revised MM5=1, M-O=2, QMSE=4, MYNN=5, P-X=7 |
| 10 | Modified PBL and surface layer? | no=0, yes=1 |
| 14–15 | Cloud Microphysics | Code as in WRF: WSM5=04, Thompson=08, Thompson+aerosol=28 |
| 16–17 | Convective scheme (D1,D2) | Code as in WRF: No=00, K-F=01, B-M=02, Grell-Devenyi=93 |
| 18–19 | SW/LW radiation | Code as in WRF: CAM=03, RRTMG=04, Fast RRTMG=24 |
| 20–21 | Separator (if need) + extra option | A, B, C, e.g. two-way nesting |

*Author contributions.* AH wrote the first draft, coordinated the sensitivity experiments and analysed some of the results. TS carried out the verification of the model results and worked on the application of the EMD. All authors participated in the design and conduction of the sensitivity experiments and in the writing and editing of the manuscript.

*Competing interests.* The authors declare no competing interests.

*Acknowledgements.* The European Commission (EC) partly funded ~~NEWA~~ the NEWA project (*NEWA- New European Wind Atlas*) through FP7 (topic FP7-ENERGY.2013.10.1.2) The authors of this paper acknowledge the support the Danish Energy Authority (EUDP 14-II, 64014-0590, Denmark); Federal Ministry for the Economic Affairs and Energy, on the basis of the decision by the German Bundestag (Germany - ref. no. 0325832A/B); Latvijas Zinatnu Akademija (Latvia); Ministerio de Economía y Competitividad (Spain -refs. no. PCIN-2014-017-C07-03, PCIN-2016-176, PCIN-2014-017-C07-04 and PCIN-2016-009); The Swedish Energy Agency (Sweden); The Scientific and Technological Research Council of Turkey (Turkey-grant number 215M386).

ANH additionally acknowledges the support of the Danish Ministry of Foreign Affairs and administered by the Danida Fellowship Centre under the project "Multiscale and Model-Chain Evaluation of Wind Atlases" (MEWA) and the ForskEL/EUDP (Denmark) project Offshore-Wake (PSO-12521/EUDP 64018-0095).

The tall mast data used for the verification has been kindly provided by the following people and organisations: Cabauw, Data provided by Cabauw Experimental Site for Atmospheric Research (Cesar), maintained by KNMI; FINO 1,2,3, German Federal Maritime And Hydrographic Agency (BSH); Ijmuiden, data from the Meteorological Mast Ijmuiden provided by Energy Research Center of the Netherlands (ECN), processed data shared by Peter Kalverla from Wageningen University, Høvsøre, Østerild, Risø, data provided by Technical University of Denmark (DTU). Most of the WRF model simulations were initialised using ERA5 data, downloaded from ECWMF and Copernicus Climate Change Service Climate Data Store.

We acknowledge PRACE for awarding us access to MareNostrum at Barcelona Supercomputing Center (BSC), Spain, without which the NEWA simulations would not have been possible. Part of the simulations were performed on the HPC Cluster EDDY at the University of Oldenburg, funded by the German Federal Ministry for Economic Affairs and Energy under grant number 0324005. This work was partially supported by the computing facilities of the Extremadura Research Centre for Advanced Technologies (CETA-CIEMAT), funded by the European Regional Development Fund (ERDF), CIEMAT and the Government of Spain. In addition, simulations carried out as part of this work also made use of the computing facilities provided by CIEMAT Computer Center.

Caroline Draxl and Gert-Jan Steeneveld are thanked for their earlier review of the WRF model sensitivity experiments. ~~Finally, we~~ We would like to thank the project and work package leaders of the NEWA project: Jakob Mann, Jake Badger, Javier Sanz Rodrigo and Julia Gottschall. Finally, we would like to thank two anonymous reviewers who took the time to carefully read, make many comments and suggestions, which greatly improved the quality of the manuscript.

## References

Anderson, J. R., Hardy, E. E., Roach, J. T., , and Witmer, R. E.: A land use and land cover classification system for use with remote sensor data, Tech. rep., United States Geological Service, https://pubs.usgs.gov/pp/0964/report.pdf, 1976.

Badger, J., Frank, H., Hahmann, A. N., and Giebel, G.: Wind-climate estimation based on mesoscale and microscale modeling: Statistical-dynamical downscaling for wind energy applications, J. Appl. Meteorol. Clim., 53, 1901–1919, https://doi.org/10.1175/JAMC-D-13-0147.1, 2014.

Benjamin, S. G., Grell, G. A., Brown, J. M., and Smirnova, T. G.: Mesoscale weather prediction with the RUC hybrid isentropic-terrain-following coordinate model., Mon. Wea. Rev., 132, 473–494, https://doi.org/10.1175/1520-0493(2004)132<0473:MWPWTR>2.0.CO;2, 2004.

Bosveld, F. C.: Cabauw In-situ Observational Program 2000 – Now: Instruments, Calibrations and Set-up, Tech. rep., KNMI, http://projects.knmi.nl/cabauw/insitu/observations/documentation/Cabauw_TR/Cabauw_TR.pdf, 2019.

Chávez-Arroyo, R., Lozano-Galiana, S., Sanz-Rodrigo, J., and Probst, O.: Statistical-dynamical downscaling of wind fields using self-organizing maps, Appl. Therm. Eng., 75, 1201–1209, https://doi.org/10.1016/j.applthermaleng.2014.03.002, 2015.

Collins, W. D., Rasch, P. J., Boville, B. A., Hack, J. J., McCaa, J. R., Williamson, D. L., Kiehl, J. T., and Briegleb, B.: Description of the NCAR Community Atmosphere Model (CAM 3.0), Tech. Rep. NCAR/TN-464+STR, Mesoscale & Microscale Meteorology Division, NCAR, USA, 2004.

Copernicus Land Monitoring Service: CORINE Land Cover, https://land.copernicus.eu/pan-european/corine-land-cover, accessed: 2019-04-15, 2019.

Danielson, J. J. and Gesch, D. B.: Global multi-resolution terrain elevation data 2010 (GMTED2010), Tech. Rep. 2011-1073, U.S. Geological Survey Open-File Report, 2011.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kallberg, P., Koehler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorolog. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Dellwik, E., Arnqvist, J., Bergström, H., Mohr, M., Söderberg, S., and Hahmann, A.: Meso-scale modeling of a forested landscape, J. Phys. Conf. Ser., 524, https://doi.org/10.1088/1742-6596/524/1/012121, 2014.

Donlon, C. J., Martin, M., Stark, J. D., Roberts-Jones, J., Fiedler, E., and Wimmer, W.: The Operational Sea Surface Temperature and Sea Ice analysis (OSTIA)., Remote Sens. Environ., 116, https://doi.org/10.1016/j.rse.2010.10.017, 2012.

Dörenkämper, M., Olsen, B. T., Witha, B., Hahmann, A. N., Davis, N., Barcons, J., Ezber, Y., García-Bustamante, E., González-Rouco, J. F., Navarro, J., Marugán, M. S., Sīle, T., Trei, W., and Žagar, M.: The Production of the New European Wind Atlas, Part 2: Production and Validation, Geosci. Model Dev. Discuss., in review, https://doi.org/10.5194/gmd-2020-23, 2020.

Draxl, C., Hahmann, A. N., Peña, A., and Giebel, G.: Evaluating winds and vertical wind shear from WRF model forecasts using seven PBL schemes, Wind Energy, 17, 39–55, https://doi.org/10.1002/we.1555, 2014.

Draxl, C., Clifton, A., Hodge, B.-M., and McCaa, J.: The Wind Integration National Dataset (WIND) Toolkit, Appl. Energ., 151, 355–366, https://doi.org/10.1016/j.apenergy.2015.03.121, 2015.

Dudhia, J.: A multi-layer soil temperature model for MM5., in: The Sixth PSU/NCAR Mesoscale Model Users' Workshop, Boulder, Colorado, USA, 1996.

Edson, J., Jampana, V., Weller, R., Bigorre, S., Plueddemann, A., Fairall, C., D. Miller, S., Mahrt, L., Vickers, D., and Hersbach, H.: On the exchange of momentum over the open ocean, J. Phys. Oceanogr., 43, 1589–1610, https://doi.org/10.1175/JPO-D-12-0173.1, 2013.

Fairall, C. W., Bradley, E. F., Hare, J. E., Grachev, A. A., and Edson, J. B. a.: Bulk parameterization of air-sea Fluxes: Updates and verification for the COARE algorithm, J. Climate, 16, 571–591, https://doi.org/10.1175/1520-0442(2003)016<0571:bpoasf>2.0.co;2, 2003.

Fernández-González, S., Martín, M. L., Merino, A., Sánchez, J. L., and Valero, F.: Uncertainty quantification and predictability of wind speed over the Iberian Peninsula, J. Geophys. Res., 122, 3877–3890, https://doi.org/10.1002/2017JD026533, 2017.

Fernández-González, S., Sastre, M., Valero, F., Merino, A., García-Ortega, E., Luis Sánchez, J., Lorenzana, J., and Martín, M. L.: Characterization of spread in a mesoscale Ensemble prediction system: Multiphysics versus Initial Conditions, Meteorol. Zeitschrift, 28, 59–67, https://doi.org/10.1127/metz/2018/0918, 2018.

Floors, R., Enevoldsen, P., Davis, N., Arnqvist, J., and Dellwik, E.: From lidar scans to roughness maps for wind resource modelling in forested areas, Wind Energy Science, 3, 353–370, https://doi.org/10.5194/wes-3-353-2018, 2018a.

Floors, R., Hahmann, A. N., and Peña, A.: Evaluating mesoscale simulations of the coastal flow using lidar measurements, J. Geophys. Res., 123, 2718–2736, https://doi.org/10.1002/2017JD027504, 2018b.

Frank, H. and Landberg, L.: Modelling the wind climate of Ireland, Bound.-Lay Meteorol., 85, 359–378, https://doi.org/10.1023/A:1000552601288, 1997.

García-Díez, M., Fernández, J., Fita, L., and Yagüe, C.: Seasonal dependence of WRF model biases and sensitivity to PBL schemes over Europe, Q. J. R. Meteorolog. Soc., 139, 501–514, https://doi.org/10.1002/qj.1976, 2013.

Gelaro, R., McCarty, W., and et al, M. J. S.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J. Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Gemmill, W., Katz, B., and Li, X.: Daily Real-Time Global Sea Surface Temperature - High Resolution Analysis at NOAA/NCEP., Office note nr. 260, 39 pp, NOAA/NWS/NCEP/MMAB, 2007.

Gómez-Navarro, J. J., Raible, C. C., and Dierer, S.: Sensitivity of the WRF model to PBL parametrizations and nesting techniques: evaluation of surface wind over complex terrain, Geosci Model Dev., 8, 3349–3363, https://doi.org/10.5194/gmd-8-3349-2015, 2015.

Grell, G. A. and Freitas, S. R.: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling, Atmos. Chem. Phys., 14, 5233–5250, https://doi.org/10.5194/acp-14-5233-2014, 2014.

Hahmann, A. N., Rostkier-Edelstein, D., Warner, T. T., Vandenberghe, F., Liu, Y., Babarsky, R., and Swerdlin, S. P.: A reanalysis system for the generation of mesoscale climatographies, J. Appl. Meteorol. Clim., 49, 954–972, https://doi.org/10.1175/2009JAMC2351.1, 2010.

Hahmann, A. N., Vincent, C. L., Peña, A., Lange, J., and Hasager, C. B.: Wind climate estimation using WRF model output: Method and model sensitivities over the sea, Int. J. Climatol., 35, 3422–3439, https://doi.org/10.1002/joc.4217, 2015.

Hahmann, A. N., Davis, N. N., Dörenkämper, M., Sīle, T., Witha, B., and Trey, W.: WRF configuration files for NEWA mesoscale ensemble and production simulations, https://doi.org/10.5281/zenodo.3709088, 2020.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G. D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., ías Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 Global Reanalysis, Q. J. R. Meteorolog. Soc., 2020, 1–51, https://doi.org/DOI: 10.1002/qj.3803, 2020.

Hong, S.-Y., Dudhia, J., and Chen, S.-H.: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation., Mon. Wea. Rev., 132, 103–120, https://doi.org/10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2, 2004.

Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Mon. Wea. Rev., 134, 2318–2341, https://doi.org/10.1175/MWR3199.1, 2006.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long–lived greenhouse gases: Calculations with the AER radiative transfer models., J. Geophys. Res., 113, D13 103., https://doi.org/10.1029/2008JD009944, 2008.

Janjic, Z. I. and Zavisa, I.: The Step–Mountain Eta Coordinate Model: Further developments of the convection, viscous sublayer, and turbulence closure schemes., Mon. Wea. Rev., 122, 927–945, 1994.

Jiménez, P., García-Bustamante, E., González-Rouco, J., Valero, F., Montávez, J., and Navarro, J.: Surface Wind Regionalization over Complex Terrain: Evaluation and Analysis of a High-Resolution WRF Simulation, J. Appl. Meteorol. Clim., 49, 268–287, https://doi.org/10.1175/2009JAMC2175.1, 2010.

Jiménez, P., González-Rouco, J., Montávez, J., Navarro, J., García-Bustamante, E., and Dudhia, J.: Analysis of the long-term surface wind variability over complex terrain using a high spatial resolution WRF simulation, Clim. Dyn., 40, 1643–1656, https://doi.org/10.1007/s00382-012-1326-z, 2013.

Jiménez, P. A., Vilà-Guerau de Arellano, J., González-Rouco, J. F., Navarro, J., Montávez, J. P., García-Bustamante, E., and Dudhia, J.: The Effect ofHeat Waves and Drought on Surface Wind Circulations in the Northeast of the Iberian Peninsula during the Summer of 2003, J. Climate, 24, 5416–5422, https://doi.org/10.1175/2011JCLI4061.1, 2011.

Jiménez, P. A., Dudhia, J., Gonzalez-Rouco, J. F., Navarro, J., Montavez, J. P., and Garcia-Bustamante, E.: A Revised Scheme for the WRF Surface Layer Formulation, Mon. Wea. Rev., 140, 898–918, https://doi.org/10.1175/MWR-D-11-00056.1, 2012.

Kain, J. S.: The Kain–Fritsch convective parameterization: An update., J. Appl. Meteorol. Clim., 43, 170–181, https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2, 2004.

Kalverla, P. C., Steeneveld, G. J., Ronda, R. J., and Holtslag, A. A.: An observational climatology of anomalous wind events at offshore meteomast IJmuiden (North Sea), J. Wind Eng. Ind. Aerodyn., 165, 86–99, https://doi.org/10.1016/j.jweia.2017.03.008, http://dx.doi.org/10.1016/j.jweia.2017.03.008, 2017.

Katragkou, E., Garciá-Diéz, M., Vautard, R., Sobolowski, S., Zanis, P., Alexandri, G., Cardoso, R. M., Colette, A., Fernandez, J., Gobiet, A., Goergen, K., Karacostas, T., Knist, S., Mayer, S., Soares, P. M., Pytharoulis, I., Tegoulias, I., Tsikerdekis, A., and Jacob, D.: Regional climate hindcast simulations within EURO-CORDEX: Evaluation of a WRF multi-physics ensemble, Geosci Model Dev., 8, 603–618, 2015.

Kruse, C., Vento, D. D., Montuoro, R., Lubin, M., and McMillan, S.: Evaluation of WRF scaling to several thousand cores on the Yellowstone supercomputer, in: Proceedings of the Front Range Consortium for Research Computing Conference, 2013.

Lee, J. A., Kolczynski, W. C., McCandless, T. C., and Haupt, S. E.: An Objective Methodology for Configuring and Down-Selecting an NWP Ensemble for Low-Level Wind Prediction, Mon. Wea. Rev., 140, 2270–2286, https://doi.org/10.1175/MWR-D-11-00065.1, 2012.

Li, D., Bou-Zeid, E., Barlage, M., Chen, F., and Smith, J. A.: Development and Evaluation of a Mosaic Approach in the WRF-Noah Framework., J. Geophys. Res., 118, 11,918–11,935, https://doi.org/10.1002/2013JD020657, 2013.

Li, S., Rupp, D. E., Hawkins, L., Mote, P. W., McNeall, D., Sparrow, S. N., Wallom, D. C. H., Betts, R. A., and Wettstein, J. J.: Reducing climate model biases by exploring parameter space with large ensembles of climate model simulations and statistical emulation, Geosci Model Dev., 12, 3017–3043, https://doi.org/10.5194/gmd-12-3017-2019, 2019.

Lucio-Eceiza, E. E., González-Rouco, J. F., Navarro, J., and Beltrami, H.: Quality Control of surface wind observations in North Eastern North America. Part I: Data Management Issues, J. Atmos. Ocean. Tech., 35, 163–182, https://doi.org/10.1175/JTECH-D-16-0204.1, 2018a.

Lucio-Eceiza, E. E., González-Rouco, J. F., Navarro, J., Beltrami, H., and Conte, J.: Quality control of surface wind observations in North Eastern North America. Part II: Measurement errors, J. Atmos. Ocean. Tech., 35, 183–205, https://doi.org/10.1175/JTECH-D-16-0205.1, 2018b.

Lupu, N., Selios, L., Warner, Z., et al.: A new measure of congruence: The Earth Mover's Distance, Political Analysis, 25, 95–113, https://doi.org/10.1017/pan.2017.2, 2017.

Mann, J., Angelou, N., Arnqvist, J., Callies, D., Cantero, E., Arroyo, R. C., Courtney, M., Cuxart, J., Dellwik, E., Gottschall, J., Ivanell, S., Kühn, P., Lea, G., Matos, J. C., Palma, J. M. L. M., Pauscher, L., Peña, A., Rodrigo, J. S., Söderberg, S., Vasiljevic, N., and Rodrigues, C. V.: Complex terrain experiments in the New European Wind Atlas, Philos. T. R. Soc. A, 375, 20160 101, https://doi.org/10.1098/rsta.2016.0101, 2017.

Mellor, G. L. and Yamada, T.: Development of a turbulence closure model for geophysical fluid problems, Rev. Geophys. and Space Phys., 20, 851–875, 1982.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. a.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, J. Geophys. Res., 102, 16 663, https://doi.org/10.1029/97JD00237, 1997.

Mooney, P. A., Broderick, C., Bruyère, C. L., Mulligan, F. J., and Prein, A. F.: Clustering of observed diurnal cycles of precipitation over the United States for evaluation of a WRF multiphysics regional climate ensemble, J. Climate, 30, 9267–9286, https://doi.org/10.1175/JCLI-D-16-0851.1, 2017.

Mortensen, N. G., Heathfield, D. N., Rathmann, O., and Nielsen, M.: Wind Atlas Analysis and Application Program: WAsP 10 Help Facility, Tech. rep., DTU Wind Energy, https://orbit.dtu.dk/files/116352660/WAsP_10_Help_Facility.pdf, 2011.

Nakanishi, M. and Niino, H.: An improved Mellor-Yamada Level-3 model: Its numerical stability and application to a regional prediction of advection fog., Bound.-Layer Meteor., 119, 397–407, https://doi.org/10.1007/s10546-005-9030-8, 2006.

Nakanishi, M. and Niino, H.: Development of an improved turbulence closure model for the atmospheric boundary layer., J. Meteor. Soc. Japan, 87, 895–912, https://doi.org/10.2151/jmsj.87.895, 2009.

Nawri, N., Petersen, G., Bjornsson, H., Hahmann, A., Jónasson, K., Hasager, C., and Clausen, N.-E.: The wind energy potential of Iceland, Renew. Energ., 69, https://doi.org/10.1016/j.renene.2014.03.040, 2014.

NCAR: NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999, https://doi.org/10.5065/D6M043C6, 2000.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., et al.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, J. Geophys. Res., 116, D12 109, https://doi.org/10.1029/2010JD015140, 2011.

Noilhan, J. and Planton, S.: A simple parameterization of land surface processes for meteorological models, Mon. Wea. Rev., 117, 536–549, https://doi.org/10.1175/1520-0493(1989)117<0536:ASPOLS>2.0.CO;2, 1989.

Olsen, B. T., Hahmann, A. N., Sempreviva, A. M., Badger, J., and Jørgensen, H. E.: An intercomparison of mesoscale models at simple sites for wind energy applications, Wind Energy, 2, 211–228, https://doi.org/10.5194/wes-2-211-2017, 2017.

Olson, J., Kenyon, J., Brown, J., Angevine, W., and Suselj, K.: Updates to the MYNN PBL and surface layer scheme for RAP/HRRR, NOAA Earth System Research Laboratory, Boulder, CO, USA, http://www2.mmm.ucar.edu/wrf/users/workshops/WS2016/oral_presentations/6.6.pdf, 2016.

Pele, O. and Werman, M.: A linear time histogram metric for improved sift matching, in: Computer Vision–ECCV 2008, pp. 495–508, Springer, 2008.

Peña, A.: Østerild: A natural laboratory for atmospheric turbulence, J. Renew. and Sust. Energ., 11, 063 302, https://doi.org/10.1063/1.5121486, 2019.

Peña, A., Floors, R., Sathe, A., Gryning, S.-E., Wagner, R., Courtney, M. S., Larsén, X. G., Hahmann, A. N., and Hasager, C. B.: Ten Years of Boundary-Layer and Wind-Power Meteorology at Høvsøre, Denmark, Boundary-Layer Meteorol., 158, 1–26, https://doi.org/10.1007/s10546-015-0079-8, 2015.

Petersen, E. L.: In search of the wind energy potential, J. Renew. and Sust. Energ., 9, 052 301, https://doi.org/10.1063/1.4999514, 2017.

Pinard, J. D. J.-p., Benoit, R., and Yu, W.: A WEST wind climate simulation of the Mountainous Yukon, Atmosphere-Ocean, 43, 259–282, https://doi.org/10.3137/ao.430306, 2005.

Pleim, J. E.: A Combined Local and Nonlocal Closure Model for the Atmospheric Boundary Layer. Part I: Model Description and Testing., J. Appl. Meteorol. Clim., 46, 1383–1395, 2007.

Poulter, B., MacBean, N., Hartley, A., and coauthors: Plant functional type classification for earth system models: results from the European Space Agency's Land Cover Climate Change Initiative, Geosci. Model Dev., 8, 2315–2328, https://doi.org/10.5194/gmd-8-2315-2015, 2015.

Rabin, J., Delon, J., and Gousseau, Y.: Circular Earth Mover's Distance for the comparison of local features, in: 2008 19th Int. Conf. Pattern Recognit., pp. 1–4, IEEE, https://doi.org/10.1109/ICPR.2008.4761372, 2008.

Refslund, J., Dellwik, E., Hahmann, A. N., Barlage, M. J., and Boegh, E.: Development of satellite green vegetation fraction time series for use in mesoscale modeling: application to the European heat wave 2006, Theor. Appl. Climatol., 117, 377–392, https://doi.org/10.1007/s00704-013-1004-z, 2014.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G.: Daily High-Resolution-Blended Analyses for Sea Surface Temperature, J. Climate, 20, 5473–5496, https://doi.org/10.1175/2007JCLI1824.1, 2007.

Reynolds, R. W., Gentemann, C. L., and Corlett, G. K.: Evaluation of AATSR and TMI Satellite SST Data, J. Climate, 23, 152–165, https://doi.org/DOI 10.1175/2009JCLI3252.1, 2010.

Rife, D. L. and Davis, C. A.: Verification of temporal variations in mesoscale numerical wind forecasts, Mon. Wea. Rev., 133, 3368–3381, http://journals.ametsoc.org/doi/abs/10.1175/MWR3052.1, 2005.

Rubner, Y., Tomasi, C., and Guibas, L. J.: The Earth Mover's Distance as a Metric for Image Retrieval, International Journal of Computer Vision, 40, 99–121, https://doi.org/10.1023/A:1026543900054, https://doi.org/10.1023/A:1026543900054, 2000.

Santos-Alamillos, F., Pozo-Vázquez, D., Ruiz-Arias, J., and Tovar-Pescador, J.: Influence of land-use misrepresentation on the accuracy of WRF wind estimates: Evaluation of GLCC and CORINE land-use maps in southern Spain, Atmos. Res., 157, 17–28, https://doi.org/10.1016/j.atmosres.2015.01.006, 2015.

Siuta, D., West, G., and Stull, R.: WRF hub-height wind forecast sensitivity to PBL scheme, grid length, and initial condition choice in complex terrain, Wea. Forecasting, 32, 493–509, https://doi.org/10.1175/WAF-D-16-0120.1, 2017.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version 3, Tech. Rep. NCAR/TN-475+STR, National Center for Atmospheric Research, 2008.

Smith, E. N., Gibbs, J. A., Fedorovich, E., and Klein, P. M.: WRF model study of the Great Plains low-level jet: Effects of grid spacing and boundary layer parameterization, J. Appl. Meteorol. Clim., 57, 2375–2397, https://doi.org/10.1175/JAMC-D-17-0361.1, 2018.

Strobach, E. and Bel, G.: Regional Decadal Climate Predictions Using an Ensemble of WRF Parameterizations Driven by the MIROC5 GCM, J. Appl. Meteorol. Clim., 58, 527–549, https://doi.org/10.1175/JAMC-D-18-0051.1, 2019.

Tammelin, B., Vihma, T., Atlaskin, E., Badger, J., Fortelius, C., Gregow, H., Horttanainen, M., Hyvönen, R., Kilpinen, J., Latikka, J., Ljungberg, K., Mortensen, N. G., Niemelä, S., Ruosteenoja, K., Salonen, K., Suomi, I., and Venäläinen, A.: Production of the Finnish Wind Atlas, Wind Energy, 16, 19–35, https://doi.org/10.1002/we.517, 2013.

Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M. A., Mitchell, K., Ek, M., Gayno, G., Wegiel, J., and Cuenca, R. H.: Implementation and verification of the unified Noah land surface model in the WRF model., in: 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, Seattle, 12-16 January 2004, AMS, 2004.

Thompson, D. R., Horstmann, J., Mouche, A., Winstead, N. S., Sterner, R., and Monaldo, F. M.: Comparison of high-resolution wind fields extracted from TerraSAR-X SAR imagery with predictions from the WRF mesoscale model, J. Geophys. Res., 117, C02 035, https://doi.org/DOI 10.1029/2011JC007526, 2012.

Troen, I. and Petersen, E. L.: European Wind Atlas, Published for the Commission of the European Communities, Directorate-General for Science, Research, and Development, Brussels, Belgium by Risø National Laboratory, https://backend.orbit.dtu.dk/ws/portalfiles/portal/112135732/European_Wind_Atlas.pdf, 1989.

Vincent, C. L. and Hahmann, A. N.: The Impact of Grid and Spectral Nudging on the Variance of the Near-Surface Wind Speed, J. Appl. Meteorol. Clim., 54, 1021–1038, https://doi.org/10.1175/JAMC-D-14-0047.1, 2015.

von Storch, H. and Zwiers, F.: Statistical Analysis in Climate Research, Cambridge University Press, 1999.

Wang, W., Dudhia, J., and Chen, M.: Application of WRF - How to get better performance, National Center for Atmospheric Research, Boulder, CO, USA, http://www2.mmm.ucar.edu/wrf/users/tutorial/201901/chen_best_practices.pdf, 2019.

Westerhellweg, A., Neumann, T., and Riedel, V.: FINO1 Mast Correction, https://pdfs.semanticscholar.org/cf85/2b7bc731b071162e537edf45f9578f4ec86e.pdf, 2012.

Wijnant, I., van Ulft, B., van Stratum, B., Barkmeijer, J., Onvlee, J., de Valk, C., Knoop, S., Kok, S., Marseille, G., Baltink, H. K., and Stepek, A.: The Dutch Offshore Wind Atlas (DOWA): Description of the dataset, Tech. Rep. TR-380, Royal Netherlands Meteorological Institute (KNMI), https://www.dutchoffshorewindatlas.nl/, 2019.

Witha, B., Hahmann, A., Sile, T., Dörenkämper, M., Ezber, Y., García-Bustamante, E., González-Rouco, J. F., Leroy, G., and Navarro, J.: WRF model sensitivity studies and specifications for the NEWA mesoscale wind atlas production runs, Tech. rep., Carl von Ossietzky University of Oldenburg, https://doi.org/10.5281/ZENODO.2682604, 2019.

Wohlfart, C., Winkler, K., Wendleder, A., and Roth, A.: TerraSAR-X and Wetlands: A Review, Remote Sensing, 10, https://doi.org/10.3390/rs10060916, 2018.

Yang, B., Qian, Y., Berg, L. K., Ma, P.-L., Wharton, S., Bulaevskaya, V., Yan, H., Hou, Z., and Shaw, W. J.: Sensitivity of turbine-height wind speeds to parameters in planetary boundary-layer and surface-layer schemes in the Weather Research and Forecasting model, Bound.-Lay Meteorol., 162, 117–142, https://doi.org/10.1007/s10546-016-0185-2, 2017.