Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-349-RC1, 2020

Geoscientific Model Development Discussions

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

Anonymous Referee #1

Received and published: 4 April 2020

General comments

This paper describes an impressive set of sensitivity experiments performed with the WRF mesoscale meteorological model, so as to obtain an optimal model configuration for the production of a New European Wind Atlas: a dataset that may well become very influential in shaping Europe's renewable energy landscape. The potential impact of this dataset justifies its documentation in the scientific literature. The vast number of combination of settings that were tested make this paper relevant and interesting for the audience of Geoscientific Model Development.

I should point out that there is significant overlap between this manuscript and an earlier technical report (https://zenodo.org/record/2682604#.XnZH1VHQg5k). Now, while I really support the early reference in anticipation of a definitive journal publication, my

Printer-friendly version

Discussion paper

impression is that the manuscript still reads like a technical report. In that respect, it doesn't help that the study is presented as a fait accompli: a mere justification of the NEWA setup that can no longer be changed. Yet whilst the production of the wind atlas has finalized, I think there's actually a lot the authors can still do to make this paper useful for the audience of GMD.

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.

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.

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).

Specific comments

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 Atlast (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."

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.

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.

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

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?

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.

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).

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?

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

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.

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.

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.

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.

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.

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.

Printer-friendly version

Discussion paper

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

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.

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.

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

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?

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 judgment’ here.

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 ...”.

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.

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.

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.

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.

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.

P20 L6: “at hub height”. Does this mean that only $\sim$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!

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.

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.

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.

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.

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?

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. . .

P22 Fig13: The y-axis is unreadable.

P24 L2-3: “In that paper we conclude that. . .” ? Better than just using ‘raw’ ERA5 data?

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.

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.

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 parameterization schemes should we abandon right away, etc.

Technical corrections

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: - P2 L6-7 rewrite “but not only” - P2 L7-8 use only on of “for example . . . to name a few” - P2 L13-15 suggest “ . . . its usefulness. It has ...” - P3 L13-16 start new sentence at “however” - P3 L10 start new sentence at “however” - P3 L16-18 move “has been reported” to

Printer-friendly version

Discussion paper

beginning: “a number of studies report . . . - P3 L19 remove comma after “cases”, suggest: “two processes with opposing effects” (remove “canceling each other out”) - P7 L12 suggest: “Small changes in wind speed are (thus) amplified when converted to power.” - P8 L18-20: suggest to split in 2 or 3 shorter sentences. Remove “the aim was”, as the next sentence also states “the objective”. - P8 L22: “or if there were regional differences” can be omitted as it is already implied by the use of “whether” - P10 L4: better to split up and rephrase, instead of using “but” twice in the same sentence. - P10 L6: this sentence can also be split in two shorter sentences. - P11 L6: Unclear, long sentence. - P22 L8: unclear sentence; a.g.l. and AGL are the same. Which figures? - P23 L18: weird use of commas around “. . . change source...” - P24 L6-7: “however . . . but ...”

Other editorial remarks: âĂć P3 L12: “A large number” or “Large numbers of” âĂć P3 L21: citation without brackets âĂć P3 L28: coastal winds? Flow is ambiguous (air or water). âĂć P4 L8: Simulations (plural), or perhaps “reference configuration”? âĂć P6 L23: remove “in” âĂć P 8 L20: remove “left” (or write “left untouched”?) âĂć P15 L2: “regional” should be “region(s)”? âĂć P17 L20: “conclusions can be drawn” âĂć P18 L1: “scheme and run” both refer to a scheme/set-up/configuration, right? âĂć P21 L21: “six” instead of 6 (in line with the surrounding text) âĂć P21 L34: rephrase “which would face” âĂć P24 L2: “wind climate” âĂć P24 L10: “best optimal” âĂć P24 L17: “observational dataset”

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