Interactive comment on “Impact of model improvements on 80-m wind speeds during the second Wind Forecast Improvement Project (WFIP2)” by Laura Bianco et al.

Anonymous Referee #1
Received and published: 28 June 2019

Review of the manuscript gmd-2019-80 Impact of model improvements on 80-m wind speeds during the second Wind Forecast Improvement Project (WFIP2) by Laura Bianco et al.

Summary
Within the context of the WFIP2 experiment, the authors evaluate the HRRR model on the performance for the 80 m wind speed. In addition they test whether a set of newly implemented physics schemes and/or increased spatial resolution improve the model performance. The evaluation covers multiple seasons, multiple starting times (Z00 and Z12) and is performed against a multiplicity of observational systems. In general an increased resolution improves the model forecast, and the experimental physics suite is only beneficial in the HRRR, but not in the NEST version. Finally the authors unravel under which types of atmospheric phenomena the experiments result in a reduced or enlarged model bias. I find this study a very thorough evaluation that clearly illustrates the challenges the field faces when comparing and improving modelling systems, i.e. against different statistical metrics for different resolutions, under contrasting scale awareness of the model etc. However, I think the paper can be strengthened with limited amount of extra work in order to become more complete in terms of model variables and in terms of setting the future research agenda for model development.
Recommendation: major revisions
We thank the Referee for the thoughtful comments. We hope we have addressed all of the Referee’s concerns and we think that our manuscript did benefit from the constructive comments made by both Referees.

Major Remarks:

1. My first concern relates to the fact that this manuscript does not describe the physical package of EXP. The authors refer to earlier papers that document these modifications. While I understand the argument of doing so, as a reader I find it usually very unattractive to first read one or two other papers to understand the current one. So I would encourage the authors to reserve some room to summarize the physical settings of EXP, so it becomes more clear to the reader what settings are underlying the bias reductions. I also think this helps the paper to generate more citations.
We thank the Referee for the suggestion. We agree that having to read another paper to understand the current one is not appealing to any reader. In light of the comment from both Referees we decided to expand section 2.2 “NWP Models”, including a list along with brief summaries of the complete set of model physical parameterizations and

relevant numerical methods targeted for development in WFIP2. We still refer to Olson et al. (2019a; 2019b), which in the meantime have been accepted for publication and are available (Olson et al. 2019a as early online releases), for details on the model configurations, but we hope this addition will give the reader all the needed tools for understanding the basic model settings used as part of this analysis.

2. Although I understand that the focus of WFIP2 is on wind energy, it would be interesting for the readership to learn to what extent the model improvements also hold for wind speeds at other heights above the surface (60 m, 100m, 120 m – hub heights are rapidly increasing). One does not need to show all graphs for all heights, buts some guidance whether improved skill for the 80-m wind is also present at other levels is interesting for the readership of the paper.
We certainly agree with the Referee on this matter. In fact, the dataset collected during WFIP2 is very rich and many other studies have been, or are being, performed to verify model improvements on other variables. While this paper was in revision other papers were accepted for publication, being submitted, or in preparation (for example a paper evaluating the models in the lowest 1 km of the atmosphere using data from scanning Doppler lidars at 3 sites is in preparation) on these other aspects and for this reason we don't think it is useful to repeat the analysis presented elsewhere in our paper. Nevertheless, in accordance with the Referee's comment, to give a wider view of the impact of the WFIP2 effort, we decided to expand Section 4 (adding subsection 4.6 "Impact of model improvements on other key meteorological variables") to summarize these other results.

3. In addition, it would be interesting to report whether improved statistics for wind also generate improved statistics for other variables as boundary-layer height, wind direction, 10-m wind speed, 2m-temperature (let's say the routine synoptic variables).Again, no additional graphs are needed, but some guidance to know whether improved80-m wind also improves or deteriorates the other variables is interesting to see the consistency of the improvements.
For boundary-layer height we are working on a separate manuscript that will focus on that aspect in particular, therefore we believe it is beyond the scope of the current work. For the other key meteorological variables mentioned by the Referee, as 10-m wind speed and 2-m temperature we did summarize the results found in Olson et al (2019) in the new subsection (4.6), as already mentioned in the answer to the comment above.

4. P6, ln 1: you suggest that the drag is too active in the revised physics. Is it possible to make this more concrete? E.g. one can discuss that this excess drag only occurs for grid cells where the modified drag scheme is active (since it switches on and off depending on the Froude number). Also if the PBL height in the model is too small, the drag has its divergence over a too shallow layer, making it too active in the atmosphere though the surface drag might be correct. In addition, it would be interesting to see whether one can distinguish whether the change in drag is due to local processes (surface drag) or modified synoptic settings induced indirectly by the drag.
There are two new sources of drag: the small-scale gravity wave drag (SSGWD) and the wind farm parameterization (WFP). The SSGWD is only active in the HRRR (for dx

> 1 km), so it does not contribute to the low near-surface wind speed biases in the HRRRNEST (dx=750 m). The WFP is active in both the HRRR and HRRRNEST. Combined, these two new sources of drag contribute to the low wind speed bias in the HRRR during the night (SSGWD is not active during the day), while the WFP can help contribute to the low wind speed bias for both the HRRR and HRRRNEST during the day or night.

The SSGWD was originally designed to only parameterize small-amplitude gravity waves, excited by rolling hills or similar types of terrain characterized by standard deviations of subgrid-scale terrain of < about 150 m. However, the original form of the SSGWD allowed the stress to be keep increasing as the standard deviation exceeded 150 m, which is common in the NorthWest US. This has been modified since the model code freeze.

Yes, low PBL height biases in the stable regime can cause excessive drag due to exaggerating the divergence of the momentum stress. To limit this, within the SSGWD only, we assume that momentum stresses decrease to zero no lower than 300 m, so the SSGWD drag is always spread over a layer at least 300 m deep. We think this is reasonable, since small-scale gravity waves may propagate into the stable layer above the model-defined PBL height and may not break until they reach the more neutral residual layer immediately above the surface stable layer. This does result in some unwanted limits in the model code, but it helps to remove excessive drag that may be caused by poorly estimated PBL heights in the stable layer.

It may be interesting to better distinguish the drag due to local (surface frictional and/or form drag) *vs* the regional or synoptically modified flows indirectly caused by the drag, but since these new forms of drag typically only directly impact the lowest 300 m and typically only combine to provide a deceleration of the low-level winds between 0.1-0.5 m s-1, we suspect that the effect on the synoptic scale is very small for forecasts less than 24 hrs in length. Also, the investigation may be contaminated by the lateral boundary conditions needed in limited area modeling. Therefore, we think this extra exercise to distinguish the drag effects from local *vs* synoptic are best suited for medium range forecasts (5-10 days) within a global modeling framework.

5. I find the paper has a rather large amount of figures, while they are not always discussed in much depth. E.g. fig 14 can be removed, including the related text on P11, ln 1-18.

According to the Referee's suggestion Fig. 14 was removed from the revised version of the manuscript. Some discussion on the behavior of the models due to the different characteristics of the cold pool events highlighted in Fig. 13 remains in the text nonetheless.

In addition to that I would encourage the authors to extend the discussion about which atmospheric conditions are responsible for the model improvement. E.g. can the bias reduction be plotted against the geowind or vs atmospheric stability?

Unfortunately, we do not have observed geostrophic wind or atmospheric stability at all the sites used in this study. Some of the sites have co-located radar wind profilers (not all, though) but the maximum height reached by this instrument is well below the geostrophic wind level. Atmospheric stability could be derived at some of the sites,

where microwave radiometers are available, but these are only 3 out of the 22 used in our study, making that variable non-representative of the entire area of interest. In any case, since Fig. 1, 2, 5, 6, 7, and 8 present the statistics as a function of the time of the day, we believe that some insight about what atmospheric conditions are responsible for the model improvements could be derived by these. Therefore, according to the Referee's suggestion we pointed to the dependence on atmospheric stability in the revised version of the manuscript, specifically, where we discussed the above-mentioned figures.

6. Although I appreciate the classification of the biases along different flow patterns, the exact definitions used to classify/categorize the flow patterns is missing in the paper. As such the reproducibility of the work is hampered.
We understand the Referee's comment on the lack of the exact definitions used to differentiate between the different flow patterns. At the beginning of the campaign several meetings were organized between the meteorologists that volunteered to participate in the weather discussions to the purpose of the creation of the Event Log. The classifications were based on the available observations, operational analysis products, HRRR forecasts, satellite images, and local radiosondes. Due to the fact that not all of these were available at all times, it was not possible to base classifications on specific thresholds and definitions. It was certainly a process that involved a certain level of subjectivity, as we already pointed out in the manuscript. Nevertheless, the process involved weekly meetings during the field study with meteorologists on the project team, many with operational forecasting experience in this geographic area, during which a consensus was reached by the team, making us confident that other meteorologists would agree with the classifications we used. Also, the Event Log is accessible to the public (available on the DAP, https://a2e.energy.gov/projects/wfip2), so the reproducibility of the work is not hampered in this sense. Some additional text was added to the revised version of the manuscript about this.

7. Methodological concern: section starting at P11, ln 31: here the bias correction is applied and then it is concluded that the skills improves further. This is logical since you just removed the bias. A better way to do this is to split the data set in two parts and determine the bias correction on the first half and evaluate it independently on the second half of the data set. I could not understand from the paper whether this procedure was followed.8.
We thank the Referee for this suggestion. According to his comment we have modified the procedure used to apply the bias correction. We now split the dataset into two parts, determine the bias correction to apply from the first part and evaluate it independently on the second half of the data set.

Finally: although I appreciate the efforts to report the model improvements and its statistical evaluation, I think the paper can be strengthened by adding a section that summarizes the future research agenda concerning surface drag, the wind speed at hub heights. This is the journal of geoscientific model development, so in my opinion it should also prioritize the research efforts of the future.

Since the model code freeze, we have prioritized three research tasks related to better simulating the low-level wind speeds: (1) the inclusion of momentum transport in the new mass-flux component of the MYNN-EDMF (already completed), (2) modifying the SSGWD to only parameterize small-amplitude gravity waves associated with subgrid-scale terrain undulations < 100 m (also completed), and (3) investigating the addition of a vertically distributed form drag as opposed to represent form drag only through the surface roughness length, which is probably only valid for dx < 1 km, where the terrain is better resolved. The impact of (1) tends to increase the near-surface wind speed in the convective boundary layer, which helps to correct the low wind speed bias we measured in WFIP2. Tasks (2) and (3) are simply meant to revise the original representation of drag in the HRRR in order to make the parameterizations more physically meaningful. All of these model components need to be investigated at a variety of model resolutions spanning dx = 1 to 10 km to ensure the model parameterizations successfully adapt in behavior to only represent the physical processes that are truly not well-resolved within the model.

Minor remarks:

P5, ln 7: when reading this I was wondering whether the statistics for other metrics behaved the same. This is dealt with later on in the paper, but perhaps it is good to announce already here that RMSE scores will be discussed later on. Just for the expectation management.
We don't look at RMSE in this study, but mostly at MAE and biases and this is pointed out in the text.

P5, ln 8: ... with SIGNIFICANTLY? smaller ...
Done.

P5, ln 11-15: this is a very long and unclear sentence.
We reworded the sentence as: "*Figure 1 can be used to examine the dependence of MAE on initialization time and forecast horizon. In particular, the Z00 MAEs are smaller than the Z12 MAE values for times soon after the Z00 initialization (for the first part of the day O lines are below X lines). In contrast the Z12 MAEs tend to be smaller than Z00 values for times soon after the Z12 initialization (for the second part of the day X lines are below O lines, except for HRRRNEST EXP), meaning that the MAE increases with the forecast horizon.*"

P7, ln 4-5: paragraph of 1 sentence, should be avoided.
Done.

P7, ln 14: cite in chronological order.
Done.

P7, ln 18: .... always positive for wind speed.
Done.

P7, ln 24: model instead of models
Done.

P10, ln 12: reword “negative blue bar”
Done. The sentence has been reworded from: “*the negative blue bar in spring and summer, visible in Fig. 9…*” to: “*blue bar in spring and summer extending toward negative values, visible in Fig. 9…*”

P10, ln 18-22: these sentences read like a figure caption, so is quite redundant
According to the Referee’s suggestion we removed the sentence “*HRRR CNT is shown in red, HRRR EXP is in blue, and observations are in black. In the lower panel, gap flow days are highlighted with the red shaded areas.*”

Figure 3: I would prefer to see this graph to be revised towards a column chart since the lines between the seasons do not say much. The statistics belong only to the season and are not connected.
According to the Referee’s suggestion Fig.3 has been modified into a bar chart.