Referee #2

The paper aims to show the modeling improvements of wind ramp events of the HRRRv4 model relative to the previous version HRRRv3. The study is well-written, highly timely, and relevant. However, the conclusions of the study, that HRRRv4 is demonstrated to generally outperform HRRRv3 for ramp events is not satisfactorily supported in the evidence provided.

I recommend major revisions to the paper before it is accepted.

We thank the Referee for the thoughtful and detailed 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. In the following text, the Referee's comments are in black and our answers are in red.

General comments

- The study does not significantly quantify the influence of its assumptions. The first assumption is that model performance at 80 m is a good proxy for performance at hub height. Showing a decent overall correlation between the two model levels for the whole period is insufficient. You should at least focus on showing it for ramp events and actual observations, even if you only have a few sites available. The second assumption is that because the spatial distribution of ramp occurrence is somewhat similar in 2020 for HRRRv3 and 2021 and 2022 for HRRRv4, the comparison in performance between these different years and model versions is warranted (implied in L220-221). You should do more to show the influence of interannual variability on the results and perhaps present the results in a way that makes it easier for the reader to convince themselves of the improvements (the dots in Fig. 5 and 6 are difficult to compare).

  To address both Referees' comment on the representativeness of 10 m wind speeds to evaluate model performances at 80 m agl, we included Appendix 1 to the revised version of the manuscript. Using the HRRR output over the 2020-2022 period, we show:

  - high correlation values (R = 0.84) between wind speeds at 10 and 80 m;
  - high correlation (R = 0.82 for up ramps and R = 0.84 for down ramps) between the total number of modeled ramps at the METAR weather stations at these 2 levels (new Fig. A1.1);
  - consistency in the normalized geographical distribution of modeled ramps between the 10 m and 80 m levels (new Fig. A1.2).
  - Also, although 80 m wind speeds are not measured in many locations compared to the availability of METAR stations, we used observations collected routinely at the Central Site of the ARM Observatory in OK to show high correlation between the 10 m level and the next few levels above it (R = 0.94 for 10 m vs 80 m wind speed and R = ~0.8 for 10 m vs 80 m wind power capacity factor) for all 3 years (new Fig. A1.3).

  Regarding inter-annual variability being a possible contribution to the skill of the model at forecasting wind ramps, we agree with the Referee's concern and we now mention this possibility in the main body of the manuscript (Section 5.2, discussion of Fig. 10) and include Appendix 2 to investigate this possibility in more detail. In Appendix 2 we show that the wind speed field output at 80 m agl of the HRRR model are similar in winter months between years 2020 and 2021, but are indeed stronger in 2022, while they are stronger in summer 2020 compared to summer months of 2021 and 2022 (new Fig. A2.1). Although there is this variability in 80 m wind speed among the years, when we look at the skill score by individual years (new Fig. A2.2), we notice that while there are some differences in skill score between years 2021 and 2020 (with the same HRRRv4 model), the skill score is still improved in both years with HRRRv4 (2021 and 2022), compared to HRRRv3 (2020). This confirms that even though inter-annual variability can impact the score of the model, HRRRv4 is still doing better than HRRRv3 as previously stated.

We also included some reasoning on the purpose/implications of our study/results in Section 3: *"Ramp events can be divided into those that occur because of the strong diurnal variability within the boundary layer, and those that are associated with meteorological phenomena such as cold fronts, gust fronts, or other changes in forcing from transient mesoscale pressure gradient fields. Although the diurnal variation of wind speeds at 10 m and at several 100 m can be out-of-phase (with 10 m wind speeds decreasing during the night time hours while at 300-400 m they may increase at night due to the low-level jet) diurnal variations at both heights are driven by surface and boundary layer fluxes and turbulent mixing. If improvements to the model's parameterization of those diurnal processes increases forecast skill at 10 m, one would expect that improvements to forecast skill would also be found at greater heights within the boundary layer."*

We agree with the Referee that Fig. 5 and 6 were difficult to interpret and we modified both of them using colorbars with appropriate ranges of variability.

- Alternative hypotheses that could explain the results (the performance improvements seen), such as natural variability, are not discussed or tested, weakening the results and conclusions drawn. Given that HRRRv3 and HRRRv4 are compared across different periods, at the least, some effort must be made to rule out natural variability as the driver of differences. L230 of the paper could indicate that interannual variability is not negligible.

  We agree with the Referees' comments and added Appendix 2 to discuss the possible impact of inter-annual variability to the results. See comments above on inter-annual variability for more details on the content of Appendix 2.

- Please sure you are following the guidelines of the journal regarding notation, dates, math symbols, etc.: https://www.wind-energy-science.net/submission.html#math, e.g., "1700 UTC" -> "17:00 UTC", avoid hyphens with abbreviated units (e.g. "10-m wind"), and many more cases.

  We tried to follow the notation guidelines in the revised version of the manuscript.

Specific comments

- L100-101: The RT&M method is so central to this study that you should spell out the details here, not simply refer to another paper

  More specifics on the way the skill score of the model at forecasting wind ramps is computed are included in the revised version of the manuscript (Section 2).

- Figure 2: I suggest indicating the study area on the map(s). In general, Fig. 2 is presented but not discussed much. Perhaps you can relate the mean and standard deviation to the number of ramp events experienced. Perhaps you could even make a ramp occurrence map.

  As suggested, we included a box indicating the study area on panels a and b of Fig. 2. Some features revealed by this figure are now discussed in the text, when discussing Fig. 2, and referred to later discussion in the manuscript. Ramp occurrence maps are already presented in Fig. 5 (model) and 6 (model/obs) for the study area.

- L169-170: Please state how the temporal interpolation was done

  We have reworded this statement in the revised version of the manuscript to: *"we have linearly interpolated the METAR observations in time to the HRRR output times"*.

- L189-190: Please explain the 3-point smoother in more detail. Is it simply the average between the two? Or something else?

  As suggested, we have provided a description of that 3-point filter in Section 3.2 (*"i.e., the model output valid at 23:00 was the weighted average of the output valid at 22:00, 23:00, and 00:00 with the two outer points having 25% weight and the central time having a 50% weight, whereas the model output valid at 00:00 was the weighted average of the output valid at 23:00, 00:00, and 01:00 with the same weighting approach"*).

- Figure 4: please add the runs for the 2021-04-07 00Z and 2021-04-11 00Z initializations to the figure to allow the reader to follow the source for the red line throughout the period

  Thanks for the suggestion. Lines relative to 2021-04-07 00Z and 2021-04-12 00Z initialization times have been added to the figure, as requested.

- Figure 5: How much of the spatial variation in ramp events is explained by the variation in mean wind speed?

  Text referring to the fact that the geographical distribution of the number of wind ramp events presented in Fig. 5 agrees with the annual wind speed geographical distribution presented in Fig. 2 has now been added to the discussion of Fig. 5.

- Figures 5 and 6: it would be helpful to show the frequency distributions of all the samples. This would also help the reader see more clearly the improvements you mention

  We have included many figures to the revised manuscript and we hope they are helping better characterize the improvements we mention.

- Figure 7: It would be valuable if you reflected on how these diurnal cycles may look different for typical hub height. For example, you mention the importance of low-level jets in the text. How would they change the picture? One could, perhaps, expect a reverse cycle at higher altitudes. Also, I would suggest using local time-of-day values or indicating the typical ranges corresponding to day- and nighttime.

  As the Referee points out, at higher heights we could indeed expect a reverse cycle in the diurnal cycle of wind speed. This consideration is particularly valid at the height of the nose of the LLJ. This has been added to the revised version of the manuscript.

  Also, sunrise and sunset times have been included in the updated version of Fig. 7, as suggested.

- “Newmann” -> “Newman” in three places on page 6

  Thanks. Corrected.

- Small suggestion: your author contributions are very short and general/vague. You can take a look at https://publications.copernicus.org/services/contributor_roles_taxonomy.html and perhaps make it more specific

  The Authors’ contribution section has been expanded as requested.