Reviewer's summary of the manuscript

In "Modeling extreme precipitation over East China with a global variable-resolution modeling framework (MPASv5.2): Impacts of resolution and physics", Zhao et al. describe how the characteristics of precipitation and vertical velocity change as resolution is increased from 60 km to 4 km and as different microphysics parameterizations are used (at high resolution only). The authors use an experimental design in which they run two quasi-uniform simulations at 60 km and 4 km resolution, and where they run several variable resolution simulations: with high resolution centered over eastern China. All simulations are short (~5 day) simulations run for a specific, weather event in summer 2012, with initial conditions provided by the GFS 1-degree forecast. By comparing with station observations of precipitation during this event, the authors show that the location and intensity of precipitation better matches observations as resolution increases to 4 km. The authors also show that the choice of microphysics parameterization has a fairly large impact on the location and intensity of precipitation.

Summary of Review

Overall, the manuscript is well-written, the evaluation methods presented are sound, the manuscript appears to fit reasonably well within the "Model Evaluation" category of GMD, and it presents results that may be of use to future users of MPASv5.2. That said, I have a few significant concerns about the manuscript: it provides minimal discussion about the physical meaning of the results, it lacks discussion of some highly relevant areas of literature, and it lacks a discussion of uncertainty (or statistical significance) in comparisons across resolutions and between simulations and observations. I don't expect that these comments will require much change to the underlying analysis, but I do think they should result in a substantial amount of new or revised text. Based on this, I am recommending that the manuscript be returned to the authors for major revisions

Major issues

Lack of discussion of physical meaning of results

Overall, the manuscript reads more like a technical report than a scientific manuscript; it focuses much more on questions of 'what' than questions of 'why'. In my opinion, this severely limits the usefulness of the paper. In its current form, I suspect that the only readers who might find the manuscript interesting would be users of the MPAS-Atmosphere model, since it essentially only focuses on describing how precipitation and vertical velocity characteristics depend on resolution and microphysics. Instead, if the manuscript had a stronger emphasis (even speculative) on why, the manuscript might be of interest to other model users facing similar questions about the effects of resolution and parameterization.

For example, in Section 3.2.2, the authors present an intriguing result: the GFS model (which are used as initial conditions!) has precipitation that is shifted far too much to the north, whereas the MPAS simulations have the rain band much closer to where it is observed. But the authors provide no speculation on why this might be, nor do they even comment that this is interesting that the MPAS model is able to 'correct' an error in the GFS starting condition. Could it be because of better-resolved topography? Is the northward propagation of the rainband perhaps less rapid in MPAS than in GFS? Are there possibly eddy-mean-flow interactions that MPAS resolves that could cause the rain band to be shifted relative to GFS?

That is just one example; this lack of exploration of 'why' is pervasive in the manuscript. A symptom of this is that almost all of the paragraphs in Section 3 have a fairly repetive structure in which they (1) introduce a new figure, (2) synthesize information contained in that figure, and (3) report some set of model performance metrics for each run (e.g., spatial correlation coefficients). That said, the authors do explore the effects of resolution on updraft velocity, which does start to get at questions of 'why', but their analysis of this is somewhat superficial, and as discussed in the section below, it misses some key literature that could enrich their analysis and discussion of this.

In summary, the authors should dig quite a bit more deeply in the analysis of their results. I would hope to see mini-hypotheses and hypothesis tests for some of the interesting intra-experimental differences that they show.

Missing discussion of key literature

The authors devote a significant portion of their analysis and discussion to the connection between vertical velocity and precipitation. This is good, but considering how significant this discussion is to the paper, the authors should discuss how these results relate to a number of recent papers on this connection.

Specifically, there are currently 3 theories in recent literature for why vertical velocity depends on resolution (with the subtext in these manuscripts that these theories can help explain the resolution dependence of precipitation):

Rauscher, S.A. et al. “A Multimodel Intercomparison of Resolution Effects on Precipitation: Simulations and Theory.” Climate Dynamics 47, no. 7–8 (October 27, 2016): 2205–18. doi:10.1007/s00382-015-2959-5.

Jeevanjee, N. “Vertical Velocity in the Gray Zone.” Journal of Advances in Modeling Earth Systems 9, no. 6 (October 2017): 2304–16. doi:10.1002/2017MS001059.

Herrington, A.R., and K.A. Reed. “An Explanation for the Sensitivity of the Mean State of the Community Atmosphere Model to Horizontal Resolution on Aquaplanets.” Journal of Climate 30, no. 13 (July 2017): 4781–97. doi:10.1175/JCLI-D-16-0069.1.

Rauscher et al. suggest that the resolution dependence results from an interaction between the constraint of fluid continuity and macro-scale turbulence. Jevanjee suggests that the the resolution dependence is related to the aspect ratio of ascending parcels, which he argues scales with resolution. Herrington and Reed suggest that the resolution dependence is related to the horizontal-wavelength-dependent growth rate of buoyancy wave instabilities.

At a minimum, this manuscript should discuss these theories, and it would be interesting if the authors provided some sort of analysis that attempts to evaluate these theories in this model. I would also suggest that the authors refer to two other relevant manuscripts: O'Brien et al. (2016) and Fildier et al. (2018), who quantitative descriptions of the connection between vertical velocity and extremes (which the authors refer to qualitatively at the end of Section 3).

O’Brien, T.A. et al. “Resolution Dependence of Precipitation Statistical Fidelity in Hindcast Simulations.” Journal of Advances in Modeling Earth Systems 8, no. 2 (June 2016): 976–90. doi:10.1002/2016MS000671.

Fildier, B. et al. “Prognostic Power of Extreme Rainfall Scaling Formulas Across Space and Time Scales.” Journal of Advances in Modeling Earth Systems 10, no. 12 (2018): 3252–67. doi:10.1029/2018MS001462.

Lack of statistics

The authors make a variety of quantitative statements comparing across simulations or between simulations and observations: e.g., "As a result, the correlation coefficients between the observations and the MPAS experiments at the resolutions of 60 km, 30 km, 16 km, and 4 km are 0.20, 0.21, 0.29, 0.50 (WSM6), and 0.42 (Thompson), respectively" (lines 422-424). However, the authors do not provide any estimates of uncertainty in these quantities, which makes it difficult to assess whether they are significant. I would expect that many of them are, but if a core goal of this paper is to assess how model skill changes with resolution, the authors should be certaint that their claims are statistically robust. I see two straight-forward ways to assess uncertainty: bootstrap confidence intervals (e.g., bootstrap sample from spatial points), or running ensembles. Ideally, the authors would run more ensemble members, but I recognize that computational constraints may prohibit that. At the very least, a bootstrap analysis would allow the authors to state the sampling uncertainty in the correlation coefficients.

Related to this, it does concern me that all of the conclusions in this manuscript are based on single-member ensembles of a single event. Would these results hold if the authors simulated another event, perhaps in another season, or even if the authors ran another ensemble member? The authors should at very least acknowledge this limitation of their study, and at best run a few additional simulations to explore whether new simulations qualitatively alter their conclusions.

Miscellaneous issues

lines 100-101: "regional models limit feedback to global scale". This is true, but is that really a good point to make in this paper? The simulations only run for ~5 days, so there is very little time for feedback--e.g., from rossby wave propagation--to feedback onto global scales. In my opinion, that makes this point a bit irrelevant, and almost misleading for this paper, since it could be read to imply that this gives these variable-resolution simulations an advantage in this experimental design over limited-area model approaches.

lines 129-131: Why were these dates chosen? Presumably it is because it is a representative, strong event; or perhaps it was chosen randomly. But a cynical version of that answer could be 'because it was a date for which the model looked very good'--I truly doubt this is case, but without any discussion of the motivation for choosing this date, a reader could wonder if this date was cherry-picked.

lines 235-241: The authors provide very little discussion on the meteorological conditions preceding the event, or of the initial condition. This limits the author's and the reader's ability to interpret differences between the simluations and observations. For example, was the Meiyu front already present and propagating in the initial condition, or did it form in the day or two preceding (presumably it was already present)?

line 246: The authors should use ERA5 instead. It covers this date, it has a significantly higher resolution, and the data are very easy to obtain either directly from ECMWF or from the NCAR RDA.