

Review of the revised version of “Evaluation of Microphysics and Boundary Layer Schemes for Simulating Extreme Rainfall Events over Saudi Arabia using WRF-ARW v4.4” by Sahu et al.

The authors put significant effort in revising the manuscript which is now substantially improved.

However, before the manuscript can be published, I have a few minor points which should be addressed:

- 1) For Fig. 1 I suggest using a different color table. Depending on the screen and printer of the reader, the reddish/orange colors can hardly be distinguished. I suggest a color table like this: https://www.ncl.ucar.edu/Document/Graphics/ColorTables/topo_15lev.shtml
- 2) Please briefly mention the advantage of the 2-way nesting approach in the manuscript. Then it is fully clear for the reader what your reasons are.
- 3) Concerning the application of ERA5 pressure level data: I fully understand that running lots of simulations is computationally expensive, but I clearly see the benefit of using ERA5 model level data for *initialization* of your 17 EREs. The ERA5 data volume for the initial conditions from ERA5 would have increased by a factor of three while there is no change in the data volume of the simulations itself. Please add a brief explanation about potential drawbacks using ERA5 data on pressure levels instead of model levels for model initialization.
- 4) Please check the grid sizes in Table 3. The numbers are incorrect.
- 5) Regarding the interpolation of the WRF simulation data to the IMERGE grid: I still see the potential for a double penalty of the model. It is not about improving your results but rather than doing it more realistically in light of the model's ability to represent features which are in the range of 3-6 times the grid distance (see e.g., Skamarock 2004: <https://journals.ametsoc.org/view/journals/mwre/132/12/mwr2830.1.xml>)
- 6) I am sorry to mention it again, but in Figure 6+7 there should be no white area for values below zero. I suggest applying a grey background and setting the minimum value to either 0.1 mm or 1.0 mm and then using white colors below 0.1 mm. E.g., 0.1 mm d⁻¹ is hard to be measured (there may also be dewfall).
- 7) Regarding your answer to my comment 65: Keep in mind that you do the surface evaluation with only nine stations over an area of approx. 3000km*2700 km and that the variables you evaluate are diagnostic variables in the model. You may consider the study of Branch et al. (2021) who evaluated WRF simulations (although with an older version of WRF) over the United Arab Emirates (which likely has a similar climate during summer) using ~50 surface stations (<https://doi.org/10.5194/gmd-14-1615-2021>). Their study shows that a wind bias is apparent in the desert.