Pointwise replies to reviewer's comments on the manuscript "Evaluating Microphysics and Boundary Layer Schemes in WRF: Assessment of 36 Scheme Combinations for 17 Major Storms in Saudi Arabia" (egusphere-2025-912)

Response to the comments of Reviewer 1

We thank the reviewers for providing such positive and detailed suggestions regarding the manuscript. We have improved our work according to their concerns.

Comment 1: Please use standard terminology (e.g., "Planetary Boundary Layer – PBL" or "convection-permitting").

Response: Thank you for the constructive suggestion. We have revised the manuscript to consistently use standard terminology, such as "Planetary Boundary Layer (PBL)" and "convection-permitting resolution." All acronyms are now defined at their first occurrence to ensure clarity for the reader.

Comment 2: The way of phrasing the aims of the study (i.e., using research questions) is not appropriate for a scientific article. In addition, some of these questions are not so relevant and could be removed (e.g., #10." How generalizable are our findings?"). The same applies to section or sub-section headings, and the Conclusions, where I would avoid using questions.

Response: Thank you for this thoughtful comment. While we recognize the traditional preference for clearly stated aims over question-based formats, using research questions is common in interdisciplinary and applied studies. After careful consideration and discussion amongst the co-authors, we believe our question-driven structure effectively conveys the study's objectives and findings, particularly given the breadth and complexity of our analysis. However, we removed the Question #10 ("How generalizable are our findings?") as per your suggestion. We remain open to reformatting the structure if the editor prefers.

Comment 3: More information should be provided on why the selected processes (PBL and cloud microphysics) were investigated. How do these parameterizations influence the simulation of precipitation? For example, elaborate on what a single and a double moment scheme is. Are there other processes relevant (e.g., convection in the coarser-resolution domain)?

Response: We appreciate the reviewer's comment. In response, we have expanded the introduction to clarify the rationale for focusing on the PBL and cloud microphysics (MP) schemes. Specifically, we added the sentence: "Two key parameterization schemes that strongly influence ERE simulations include the Planetary Boundary Layer (PBL) and cloud microphysics

(MP) schemes." We then included two dedicated paragraphs: one describing the role of PBL schemes in modulating turbulent mixing, boundary-layer growth, and the initiation of convection; the other describing how MP schemes govern hydrometeor development, cloud phase transitions, and precipitation intensity. The MP section also explains single- and double-moment schemes, including their implications for predicting hydrometeor number concentrations and mass. Each paragraph contains citations to relevant literature.

We acknowledge that other processes — such as cumulus parameterizations in the coarser domain, radiation schemes, and land surface interactions — also affect precipitation simulation. However, we limited the present assessment to PBL and MP schemes because they exert the most direct control on convective processes and cloud microphysics at convection-permitting scales, as well as to limit computational and storage demands. We have clarified this scope limitation in the conclusion, adding: "To further advance ERE simulation fidelity, future work should extend beyond PBL and MP schemes to systematically evaluate the impact of land surface schemes, radiation parameterizations, and data assimilation techniques." (Lines 40-49, 51-58)

Comment 4: The selection and definition of extreme cases is problematic, since for many events there is unknown information on the observed precipitation amounts (Table 2). I recommend including an additional column which will show the IMERG nearest grid-cell precipitation.

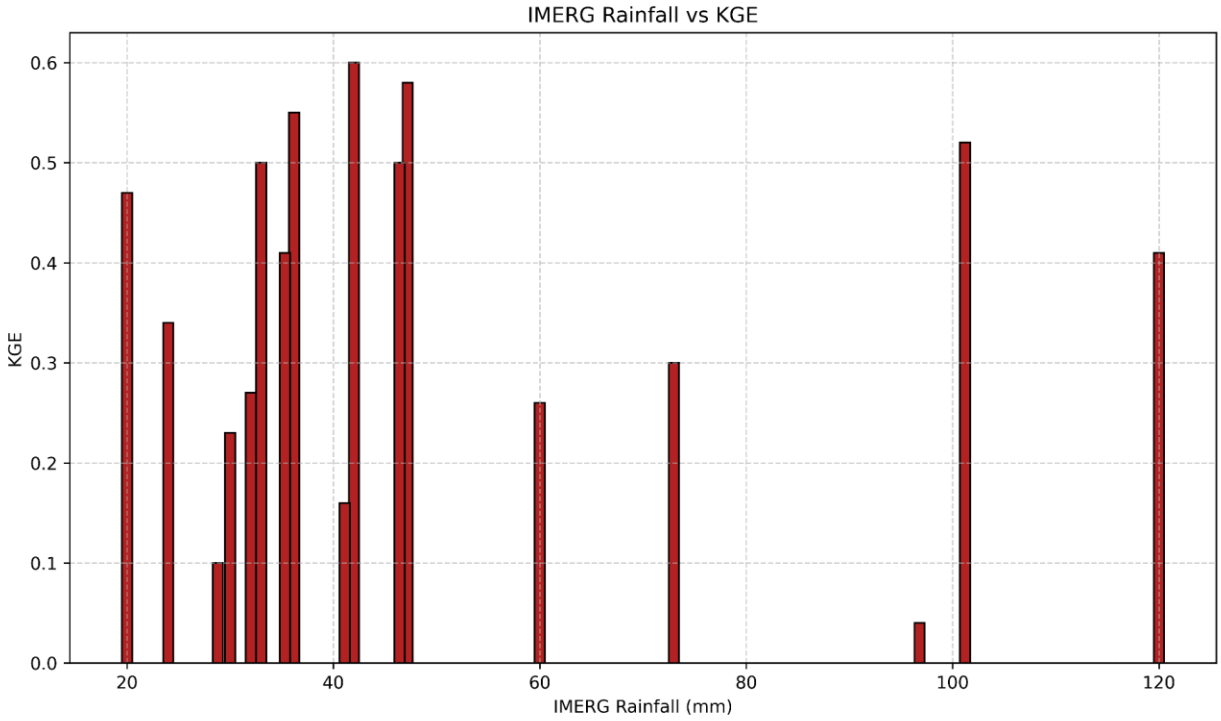
Response: Thank you for your valuable comment. We have added a new column with IMERG rainfall amounts for the events.

Comment 5: It is also unclear if any events lasted more than one day, and if yes, how were these events treated in the analysis? Is one day of spin-up time enough for these runs?

Response: Thank you for raising this critical point. In all cases, the large majority of rainfall occurred within one day. All model simulations were conducted for 84 hours, including a 48-hour spin-up period to ensure model stability and reduce initialization biases. The analysis was focused on a 24-hour window corresponding to the peak rainfall period of each extreme event (Table 2). Our study specifically targets short-duration, event-based simulations of extreme rainfall. In such cases, the primary drivers are typically large-scale atmospheric instabilities and moisture advection rather than slower processes like land–surface interactions. Consequently, a 48-hour spin-up period is sufficient to allow the model to dynamically and thermodynamically adjust to the initial and boundary conditions. This clarification has been added to the revised Data and Methods section (Line 139-145).

Comment 6: For extracting the overall statistics, all events were weighted equally. However, in the interpretation of results, it would have been useful to differentiate, for example, between the most and the less extreme events, or between events affecting different parts of the Arabian Peninsula.

Response: Thank you for this insightful suggestion. We fully agree that distinguishing between more and less extreme events, as well as regional variability, can offer valuable insights. However, in our case, we found no systematic dependence of KGE values on the rainfall intensity of the events (see the figure below). There is no discernible trend or correlation between IMERG rainfall amounts and the corresponding KGE values, suggesting that model performance does not scale with event intensity. Based on this finding, we consider that giving equal weight to all events is a reasonable and justified approach in our statistical summaries. Additionally, stratifying the results further would make an already complicated analysis even more, hindering interpretation and presentation.



Comment 7: More information on the interpretation of KGE should be provided in Section 3.5. Some references to other studies that use KGE in a spatial context could also be added. Moreover, I strongly recommend using additional evaluation metrics and not relying only on KGE for your conclusions.

Response: Thank you for the suggestion. We have clarified in Section 3.5 that KGE is an aggregate metric incorporating correlation, bias ratio, and variability ratio. To maintain clarity and avoid overwhelming the reader, we chose not to include additional performance metrics. However, we have added the following sentence to support our approach and provide references for spatial applications of KGE:

"The KGE is an aggregate performance metric that integrates correlation, bias ratio, and variability ratio into a single score, providing a holistic assessment of model performance. While additional metrics could be computed, including too many would risk overwhelming the interpretation. Several studies have successfully used KGE for spatial performance assessment of hydrometeorological models (e.g., Gupta et al., 2009; Patil and Stieglitz, 2014; Beck et al., 2019; Nguyen et al., 2022; Tudaji et al., 2025), supporting its application in our spatial analysis." (Line 148-151)

Comment 8: Extensive parts of Section 4 are not results (e.g., L148-161, L217-222, L275-280, L326-332). Please move this and other non-results material to the introduction, data or discussion sections, if relevant.

Response: Thank you for your observation. We would like to clarify that Section 4 is the Results and Discussion section. Accordingly, besides presenting the results, we provided interpretation, compared our results to other studies, and answered the questions posed in each subsection. Nevertheless, we have carefully reviewed Section 4 and agree that Lines 148–161 contain background context and methodological details. These have now been relocated to the introduction (Line 43-49). Regarding Lines 217–222, 275–280, and 326–332, we respectfully retain these in their current positions, as these paragraphs serve as essential contextual discussion that supports the interpretation of findings.

Comment 9: The approach described in lines 251-257 should be presented in more detail in the Methods section.

Response: Thank you for pointing this out. We agreed and have expanded the section as follows:

"Additionally, to determine whether the performance is significantly different between scheme combinations, we calculated ΔKGE scores by subtracting the mean KGE across events from the KGE values, thereby eliminating systematic differences in scores among events. We then tested whether the distributions of ΔKGE values for different scheme combinations are statistically similar or different using pairwise independent t-tests." (Line 166-169)

Comment 10: Figures 6 and 7 should be merged to facilitate the comparison between observations and simulated rainfall. Please be consistent in the date format (panel titles).

Response: Thank you for the suggestion. We strongly considered merging Figures 6 and 7 to facilitate direct comparison; however, doing so would result in 34 panels, which we believe would compromise clarity and interpretability. Instead, we have ensured that the date format is now consistent across all panel titles to improve the consistency and facilitate easier comparison between the figures.

Comment 11: Sections 4.9 and 4.10 are definitely not results material. Please move to other more relevant section(s).

Response: Thank you for your feedback. Our paper includes a merged Results and Discussion section for improved readability. Sections 4.9 and 4.10 were originally intended to revisit the research questions and synthesize key insights. However, we acknowledge these were more interpretative in nature. We have retained Section 4.9, as it revisits the research questions and synthesizes key findings in a way that supports the overall coherence of the paper. Section 4.10 has been removed as suggested.