Response to Reviewer 2's comments

[General] *Cloud microphysical processes are key components in parameterizing precipitation in numerical models yet large uncertainties remain between different autoconversion schemes. By combining four autoconversion rates schemes through a weight mean approach, the authors propose an ensemble scheme to try to avoid limitations of individual scheme. The ensemble scheme is then incorporated into the Thompson scheme to simulate an extreme rainfall event over Southern China. The rainfall extreme, distribution (both temporal and spatial) and hydrometer content are then compared with simulation with the Berry and Reinhardt (1974) scheme. Results show improvements in the timing and space of rainfall peak. This manuscript is well written, and the topic of this manuscript fits the scope of GMD. I recommend acceptance for publication after returning to the authors for minor revision.*

Response: Thank you very much for agreeing with us to the intention of this manuscript. We appreciate you for providing valuable comments and constructive remarks, which have helped improve our manuscript significantly

[Major]

The authors choose to compare simulation from EN with that from BR, I understand that it is partially because BR is used in the original Thompson scheme, but some results are kind of expected from Figure 2, for example, delayed rainfall peak. Did you compare the EN results with simulation using LD scheme?

Response: Yes. As has been addressed above, it is convenient to conduct a simulation with any of the above-listed schemes alone. In total, five experiments were carried out with the EN，KS，BR，KK, and LD schemes. The results indicate that the EN scheme provides better simulations than those treated by using any single scheme alone in terms of accumulated rainfall and extreme hourly rainfall rate.

Figure R1 compares the spatial distribution of 18-h simulated total rainfall from the simulations with the EN，KS，BR，KK and LD schemes to the observed. Generally speaking, all the schemes are able to capture the main characteristics of the extreme rainfall event. One can see that the simulated rainfall amount compares favorably to the observed both at HS and JL, although the JL storm has a 10-15 km eastward location shift. Comparatively speaking, the EN and BR schemes performed better than others. The two centralized rainfall cores over HS and JL were successfully captured by the EN and BR schemes, with the simulated heaviest rainfall amount of 537 mm and 569 mm, respectively (Fig. 1b,d). As for the EN scheme (Fig. R1b), the simulated 18-h total rainfalls were 320 mm and 537 mm over HS and JL, respectively, which was close to the observations of 341 mm and 542 mm (Fig. R1a). Similarly, the BR scheme performed similar to the EN scheme, with the maximum rainfall of 347 mm and 569 mm over Huashan and Jiulong regions, respectively (Fig. R1d). One unique feature of the observations was the rapid increase in the hourly rainfall rate. The rainfall produced by the EN scheme peaked within 2 h while the BR scheme peaked over a period of 4 h. Both the simulated rainfall rates decrease for several hours. Generally speaking, the EN scheme performed much closer to the observed, compared to that of the BR scheme. Note that the longer heavy rainfall period from the BR scheme contributed partially to the over-prediction of the 18-h accumulated rainfall. In terms of the temporal evolution of radar reflectivity, one can find that the Jiulong storm simulated with the EN scheme (Fig. 5f) developed more rapidly than that from the BR scheme, almost 1 h earlier than the latter (Fig. 5i). This was consistent with the timing lag in the hourly extreme rainfall production (Fig. 4).

The heavy rainfall amounts over Jiulong region were underestimated by the KS, KK, and LD schemes, with the heaviest rainfall amounts of 434 mm, 463 mm, and 473 mm, respectively (Fig. R1c,e,f). Note that the simulated heaviest over Huashan region were comparative among each other.

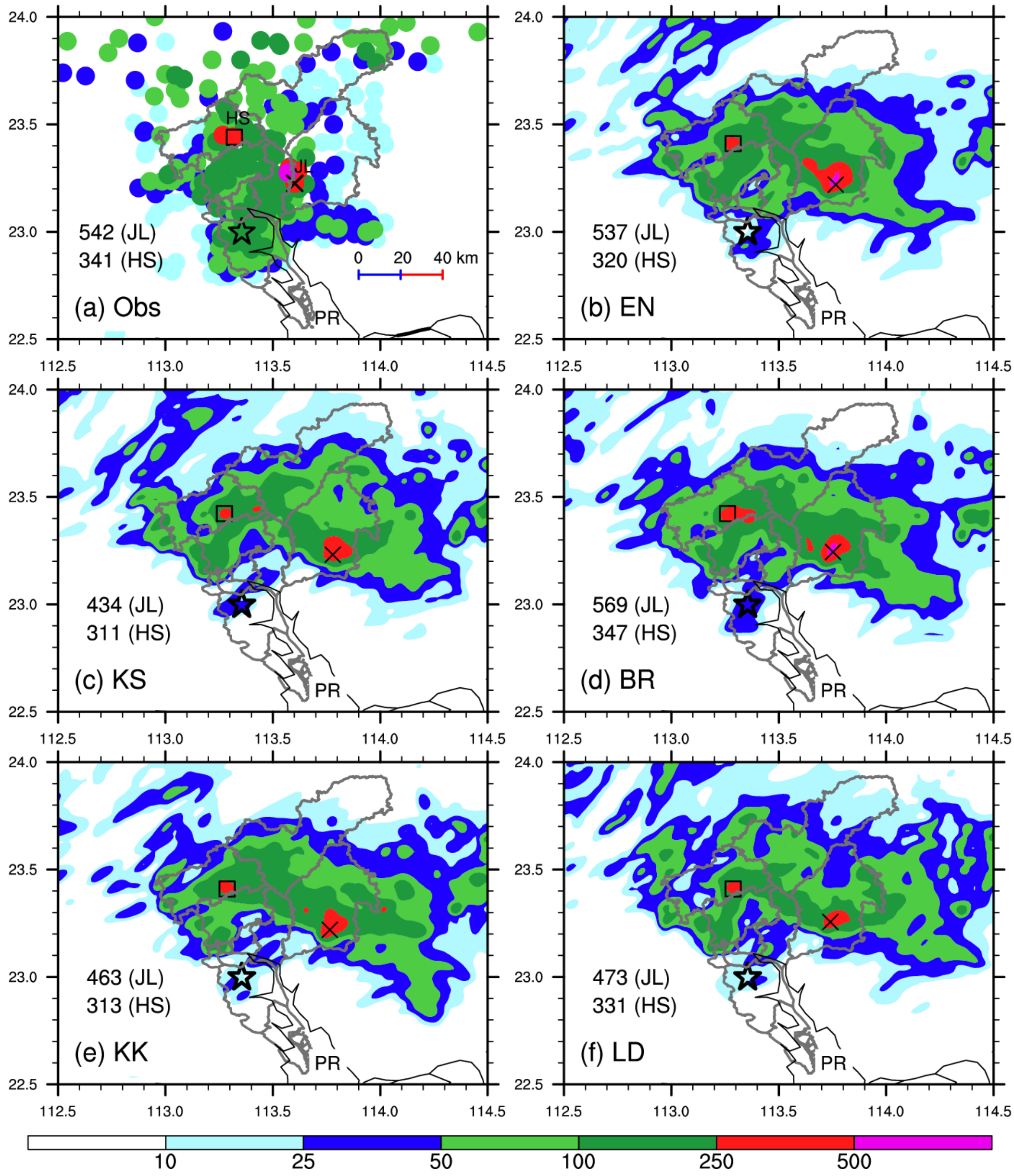

Fig. R1 Spatial distribution of the 18-h accumulated rainfall during the period from 2000 Beijing standard time (BST, BST = UTC + 8) 6 May to 1400 BST 7 May 2017. (a) rain gauge observations, and (b-f) simulations with various autoconversion schemes during the. A cross sign (×) and a square sign (□) denote the locations where maximum hourly rainfall rates were (a) observed or (b-f) simulated near Jiulong (JL) and Huashan(HS), respectively. The values marked with JL and HS indicate the 18-h maximum accumulated rainfall amounts near the JL and HS, respectively. A star indicates the city center of Guangzhou, and the Pearl River is marked by PR.

I appreciate the efforts of combining different schemes, but the manuscript lacks descriptions and recommendations on how to adjust the weights in the EN when simulating clouds in different synoptic systems, for example, continental deep convection vs maritime drizzling stratocumulus. As the authors stated in Section 2 that each of the schemes spatializes in certain conditions. In the case demonstration, if you adjust the weights to giving more weightings to schemes that are more suitable for continental deep convection, will the results be closer to observations? It might be too much work to add in this manuscript, but the EN scheme will be more practically valuable if the authors can propose a recommending framework to adjust the weights for different types of clouds.

Response: Thanks for your constructive comment. Adjusting the weights in the EN scheme should give better results for different synoptic systems. At present, it is troublesome to provide recommended weights for different synoptic systems without a large number of tests and verification for specified weather conditions. In this study, we focused on the EN approach and provided a flexible adjustment interface for different aims. Users can adjust the weights according to their objectives, even easily planting new members into the EN scheme. In order to help users understand the options, a detailed description of the selected autoconversion schemes (i.e., KE, BR, KK, and LD) has been added in the revised manuscript. Keeping your suggestions in mind, a recommending framework to adjust the weights for different types of clouds will be updated with the source codes on Zenodo (https://doi.org/10.5281/zenodo.5052639) after detailed experiments in the future.

[Minor]

Line 99-100: please rephase this sentence. Do you mean the Cotton (1972) scheme results in the peak cloud water content occur the earliest time, at the lowest cloud attitude but has the lowest value as compared with other schemes?

Response: Thank you very much for pointing this out. We have made revisions accordingly.

Line 119: remove are

Response: Thank you very much for the reminder. Removed.

Line 222-230: I do not get how the ensemble scheme can represent subgrid-scale cloud processes with integrating one or more of the schemes over any assumed CWC or Nc distributions like in Griffin and Larson, 2013. Any one of the four schemes itself cannot represent subgrid-scale processes.

Response: Not really. To the best of our knowledge, each individual scheme has its own advantages and disadvantages, and there is no one scheme able to provide good results at all times. For example, the LD scheme considering spectral dispersion was more reliable for improving the understanding of the aerosol indirect effects, and the KK scheme aimed at large-eddy simulation (LES). With the development of the variable resolution models, it is flexible to represent cloud processes consistently across all model scales under various conditions. Depending on grid distance, one or more schemes can be used independently in a variable resolution model. To avoid misunderstanding, the word "*subgrid-scale*" has been removed.

Line 288: …it is convenient to *conduct* a launch simulation…

Response: Thanks for your kind reminders. We revised the sentence as follows: "it is convenient to conduct a simulation…"

Line 321: what is 'ER'? please elaborate when you first introduce an abbreviation.

Response: ER denotes extreme rainfall. Corrected.

Figure 7: is there radar observations at Jiulong site to compare reflectivity in observation and simulations? Does the observed maximum reflectivity extend to the surface?

Response: The observed composite radar reflectivity was integrated by combining four individual radar observations at Guangzhou and its surroundings. Yes, the observed maximum stretched to the ground. Please refer to our previous observational analysis for detailed radar reflectivity vertical structures of the extreme rainfall, which is given in Li et al. (2020).

Li, M., Y. Luo, D. L. Zhang, M. Chen, C. Wu, J. Yin, and R. Ma, 2021: Analysis of a Record-Breaking Rainfall Event Associated With a Monsoon Coastal Megacity of South China Using Multisource Data. IEEE Transactions on Geoscience and Remote Sensing, 59, 6404-6414, doi:10.1109/TGRS.2020.3029831.

We appreciate you very much for your positive and constructive comments and suggestions on our manuscript, which are valuable in improving the quality of our manuscript.