This is an interesting study giving a statistical evaluation of different microphysical schemes using real radar observations on 30 convective cases.

The first part of the results dedicated to the analysis of heavy rain is very convincing. The two ways of sorting the results as a function of reflectivity first and of rain content next is nice and helps showing the effect of the understimation or overestimation (depending on the scheme) of the number of large drops in the PSD.

Regarding the part with hail and graupel statistics, I don't always agree with the interpretation of the results.

Thank you for this review. Our point-by-point review is in blue. The changes made to the manuscript are highlighted in red. A marked up-version showing all changes to the manuscript is provided along with the revised manuscript.

l 317: "large graupel and hail produce similar radar signals"

It is true only if graupel and hail are modeled with the same characteristics (PSD and density for Zh). Depending of the density options for graupel in the microphysics scheme you are evaluating, you could underestimate the maximum possible reflectivities that would be reached if explicit with hail (which is denser than graupel) was modeled in your schemes.

There are several instances where the classification algorithm identifies hail even though the original model does not simulate it. We still believe that these "false" identifications are primarily due to large graupel particles, since they are similar in shape to hail and therefore have similar differential reflectivity, and large graupel particles can also produce large reflectivities. However, it is true that hail would produce even higher reflectivities due to its higher density. We did, in fact, change the 'graupel/hail' analysis to a general 'ice' analysis, to allow a fairer comparison between the P3 scheme and the other schemes. Therefore, this sentence is not required anymore and we deleted it completely.

l 327: "Since none of the simulations, regardless of the cloud microphyics scheme, were able to reproduce these extreme events, we do not believe that this is related to the microphysics scheme, but rather a consequence of the model grid resolution."

The resolution very probably plays a role but again, you can't simulate the extreme reflectivities due to hail (in your observation) while you don't explicitly have hail in your model (and the corresponding options in the forward operator).

It is true that we underestimate the maximum possible reflectivity if we do not explicitly calculate hail (which has a higher density than graupel). This has not been taken into account so far. We have added a part that takes into account the influence of density on the simulated reflectivity. The exact wording can be found in our response to the next comment.

L 363: “The SBM scheme, on the other hand, is again likely missing the larger particles, since a large mass of graupel particles is generated, but this does not translate into high reflectivities”

Again, could this be also due to a different graupel density in this scheme compared to others ?

More information about the differences in the density of graupel / rimed fraction between the different schemes (and compared to typical hail density) should be included in the paper, either to evaluate if this could have an effect or not.

The difference in graupel density between the SBM scheme and most of the other schemes is not drastically different. The Morrison and SBM scheme use 400 kg m^{-3} and the two Thompson schemes use 500 kg m^{-3} as a constant graupel density. The P3 scheme uses a varying ice density, that can reach up to 900 kg m^{-3}. We don’t think that the graupel density of the SBM scheme is the main reason for the low reflectivity produced, given that the density is not drastically different to most other bulk schemes. However, we understand that the density is playing a major role in general when simulating reflectivity, and should be discussed more in detail. Especially for the P3 scheme, it was not yet acknowledged that the more flexible density assumptions (reaching up to 900 kg m^{-3}, i.e., hail-like particles) might produce more realistic reflectivities compared to the observations.

We acknowledge the density influence now at multiple instances:

Section 5.1:

> *(2) Particle density: The particle density strongly influences the reflectivity. All schemes except the P3 scheme consider graupel particles with constant density of 400 - 500 kg m^{-3}, and do not explicitly calculate hail. Hail particles, however, are typically much denser than graupel. This means, if hail events are observed, the high hail density can lead to high observed reflectivities that cannot be reproduced by the models due to the lower assumed particle density. Only the P3 scheme has a more flexible approach that allows varying ice particle density reaching up to 900 kg m^{-3}.*

Section 5.1:

> *The P3 is also the only scheme that allows ice particles to reach densities up to 900 kg m^{-3}, i.e., to simulate hail-like particles.*

Section 5.1:

> *However, graupel density assumptions could also play a role. Both, Morrison and SBM assume a graupel density of 400 kg m^{-3} which is slightly lower than assumed by the Thompson schemes with 500 kg m^{-3}.*

Section 6:

This might be related 1) to a resolution problem, 2) to density assumptions that are not representative for high density hail-like particles or 3) the absence of partially melted particles in the simulations.