Review of the paper "Can models robustly represent aerosol–convection interactions if their cloud microphysics is uncertain?", authored by B. White, E. Gryspeerdt, P. Stier, H. Morrison, and G. Thompson.

The authors have drawn a pretty grim picture of the situation with the description of microphysics in cloud resolving models using bulk parameterization schemes. Two known bulk parameterization schemes referred to as "MORR" (Morrison and Milbrandt (2011) and "THOM" (Thompson et al. (2004, 2008)) were tested by simulation of 3 case studies with different type of convection: from shallow convection to a supercell storm. The simulations were performed at a priory given droplet concentrations of 100 cm-3, 250 cm-3 and 2500 cm-3. A dramatic difference between parameters of simulated clouds (cloudiness, simulated precipitation, etc.) and observations is demonstrated. A huge difference in cloud microphysical structure simulated by these two schemes is reported. Both bulk schemes turned out to be insensitive to droplet concentration. So, the difference in results are related to the differences in the bulk-parameterization schemes.

My particular comments and remarks are the following.

1. The finding about high diversity of cloud microstructure and precipitation produced by different bulk schemes is not new. Any intercomparison study shows such diversity. The insensitivity of most bulk schemes to aerosols (and to droplet concentration) also well known. For instance, application of different bulk -parameterization schemes to simulation of hurricane Irene (2011), including "MORR" and "THOM", led to a TC with maximum wind varying from 30 m/s to 70 m/s (Khain et al., 2015; Khain et al. 2016-Atmospheric Research 167, 129–145). In these studies insensitivity of bulk schemes to aerosols was also demonstrated and an possible explanation of such insensitivity is presented.

We agree with Reviewer 1 that it has been shown that different microphysics schemes produce a high diversity of cloud development and precipitation. We have significantly revised our introduction to show this more explicitly, and in doing so have included the references and possible explanations suggested by the Reviewer. This is in P1 L20 through P8 L4 of our revised manuscript.

We certainly do not intend to claim this in itself as a new result. However, we confirm this in a single modelling framework, for three different cloud and environment types, in three different simulation types (one performed with meteorological data, one idealised supercell using open boundaries, one warm-phase cumulus case with periodic boundaries). This is necessary to show before presenting our main results: (a) that the largest source of uncertainty (or variability) between simulations is in the choice of microphysics scheme, and that any aerosol effect is secondary, and (b) that the response of the hydrometeors to CDNC perturbations differs strongly not just between microphysics schemes but also that the inter-scheme variability differs between cases of convection

Although they do not always respond to aerosol as strongly as bin schemes, bulk schemes have been shown to be sensitive to aerosol, e.g. Morrison & Grabowski 2011, who found an ice-phase response to aerosol. Indeed, a similar mechanism was later confirmed in bin-scheme simulations by Fan et al. 2013. Further, Kalina et al. 2014 found that autoconversion of cloud water to rain decreased under polluted conditions and, subsequently, near-surface rain and hail particles increased in size due to enhanced collection of cloud droplets. Although the sensitivity of the bulk schemes used in the current study to perturbations in cloud droplet numbers was secondary to the sensitivity in cloud and precipitation development dependent on the choice of bulk scheme, we nevertheless feel this is a significant result to present in the context of multiple cloud types and types of simulation, given that bulk microphysics schemes remain in wide use.

Futher, in idealised supercell simulations using the WRF-SBM, we find the same order of magnitude of aerosol effect as in our bulk scheme simulations, and uncertainty due to microphysics scheme still dominates aerosol effects (Figure S4 in the Supplement).

We present in Figure S1 in the Supplement joint histograms for our Congo basin simulations of cloud top height in convective updraughts (columns identified with vertical velocities greater than 1 ms^{-1}) and the radius of the updraughts, as identified by running a connected-components labelling algorithm on the field of identified updraughts. Figure S1 shows that the most significant dynamical difference comes from choice of microphysics scheme: the Morrison scheme has a tendency towards higher frequencies of wider updraught radii with higher cloud tops than the Thompson scheme.

We note that the updraught dynamics in the Congo simulations respond very differently to aerosol. Figure R2 shows the difference between the joint histograms for the polluted and pristine cases for each microphysics scheme, respectively. While MORR shows little consistent aerosol response, THOM shows consistently lower frequencies of occurrence of all updraught radii and a reduced frequency of occurrence of the highest updraught tops under polluted conditions, with an increased frequency of occurrence of small updraught radii with lower cloud tops. Therefore, a consistent aerosol response is observed in THOM, resulting in smaller and lower convective updraughts (i.e. weakened convection under polluted conditions). Interestingly, both of these effects contradict the findings of Morrison & Grabowski 2011, who found an ice-phase response to aerosol in which cloud top heights and anvil ice mixing ratios increase under polluted conditions due to increased freezing of larger numbers of cloud droplets and subsequent higher ice particle concentrations with smaller sizes and reduced fall speeds. However, we note again that we consider different values of CDNC / CCN (and responses may be nonmonotonic, Kalina et al. 2014), and different case of convection (indeed, our 10-day Congo simulation covers many convective lifecycles). This combined with the findings already presented in our study, leads us to suggest that it is not certain that a consistent response in a different case of convection and with different CDNC values would be expected.

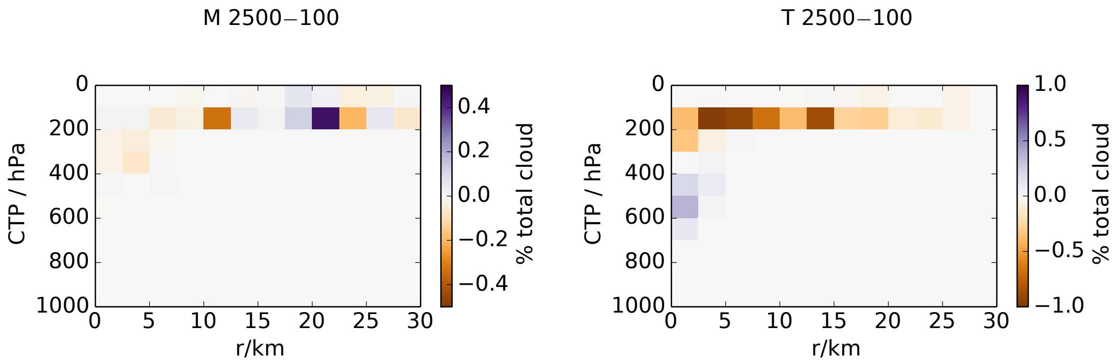

Figure R2: *Congo case: difference in joint histograms of updraught radius and cloud top pressure at the top of the cloudy updraughts between the polluted and pristine cases for MORR (left) and THOM (right).*

We have included Figure R2 in the revised paper and the relevant discussion Is on P11 L11-30.

2. The finding that description of the autoconversion is of crucial importance for correct simulation of cloud microphysics is also not new. For instance, Gilmore and Straka (2008) showed that the most formulae for autoconversion used in bulk schemes are applicable to the initial stage of the first raindrop formation only, and that the rates predicted by these formulae differ by orders of magnitude. There are many other studies showed dramatic sensitivity of cloud microstructure to the scheme of autoconversion. So, the new information in the paper is that these conclusions are confirmed in investigation of these two particular bulk schemes.
We thank Reviewer 1 for the Gilmore & Straka reference which helps to give context to our findings. We have included this reference in our Results (P18 L10) and Conclusions (P25 L24) section.

We do not claim our finding that cloud structure is sensitive to autoconversion representation to be a new result. Rather, we find that autoconversion is sufficient to explain some of the key differences between the clouds simulated by the two bulk microphysics schemes, namely those in the liquid phase. Further, we show that aerosol effects are dominated by the response of the autoconversion process for this case. Li et al. (2015, JAS) also showed that a slower autoconversion process along with a stronger accretion process explains the Morrison scheme's higher cloud fraction than the Thompson scheme for a similar rain mixing ratio.

This is in part a finding in itself that addresses Reviewer 1's comment 4. It is because we were trying to understand which processes were significant to the differences in cloud microstructure in the context of comparing these two schemes that we investigated the autoconversion representation. Indeed, we cannot say in the cases we consider which is the more 'correct' representation. We rather highlight that some of the differences we observe between the schemes, in some of the types of convection, can be almost entirely attributed to the representation of autoconversion, and that differences in other processes between the two schemes are secondary.

Further, our tests highlighted the difference between the MORR and THOM schemes in the interplay and relative importance of the autoconversion and accretion processes: the Thompson scheme can produce surface rain from autoconversion alone (although two orders of magnitude less than when rain can also accrete cloud water), showing that autoconversion acts almost like a 'seed' for rain production in this scheme, after which accretion takes over the rain production process. However, in the Morrison scheme, when autoconversion is allowed to occur but accretion of cloud water by rain is prevented, precipitation shuts down, thus indicating that both autoconversion and accretion are necessary processes for warm rain production in MORR.

This is discussed in our Results section.

3. The authors illustrate vertical profiles of mass contents of different hydrometeors averaged over the entire computational area. As a result, all information concerning the microphysical structure of clouds simulated by different schemes turns out to be lost, at least for specialists in cloud microphysics. It is necessary to present vertical profiles of maximum values of the mass contents. The profiles of cloud averaged values would be also useful. These figures should be accompanied by corresponding comments and analysis.
It is important to include the profiles averaged over the entire domain, because (at least for larger-scale impacts) this illustrates the difference in the bulk properties and accounts for differences in total cloud cover, which in turn has implications for radiative effects and

feedbacks. This is especially important in our Congo simulations, where differences in the ice-phase microphysics between the schemes lead to large differences in the ice cloud fraction, which will in turn have a significant radiative impact.

We agree with Reviewer 1 that it would also be useful to see the mean in-cloud properties, and changes thereof. We thus now include condensate-averaged profiles for each hydrometeor type in order to identify not only the bulk impact of changes in CDNC in each simulation but also the bulk properties of the cloud in each convective study. We note that in the idealised supercell case and the RICO LES case there is little difference between the domain-mean profiles and the hydrometeor class-mean profiles.

Comments and analysis of these new profiles have been included in our revised manuscript. The relevant Figures are Figures 6,7,8,9,11,12.

An example of the condensed-point average profiles is provided in our response to Reviewer 2's comment 12.

However, we do not agree that showing domain-maximum values is useful, because the maximum values of the mass contents represent just one point in the entire domain and can lead to improper conclusions, especially in our Congo simulation where multiple cloud types exist in the same domain but do not necessarily interact.

4. The lack of physical interpretation of results is another drawback of the study. For instance, Fig. 6 and Fig. 7 show that in simulations CONGO - T250, T-2500 maximum cloud water content (i.e. small cloud droplets) is located at the surface. One gets the impression that the entire boundary layer is filled with tiny droplets (but not with raindrops). What can be physical mechanisms leading to this very strange effect? The comment concerning the lack of interpretation is related to most figures.

The low cloud extending down to near-surface regions is indeed a strange effect. However, it is produced consistently in all simulations (including the simulation performed with WRF SBM microphysics, see Figures S2 and S3 in the Supplement now included with this response). We note also that in our WRF-SBM simulation (Figures S1 and S2 in the Supplement) the warm cloud mass has significantly greater domain-mean values than the M250 case. We believe this difference is because, as we have shown (our Figure 16), the cloud mass in the M250 case is removed through autoconversion to rain, whereas this does not occur in the T250 case (or indeed the SBM). When the THOM autoconversion is used in the equivalent MORR simulations (or when autoconversion is turned off completely in MORR), the cloud droplets persist as in THOM. Although we have not tested the equivalent process in the SBM simulation presented in this response, we suggest that lack of autoconversion (or, stronger autoconversion than THOM but weaker than that in MORR) is responsible for the persistence of the cloud droplets in THOM. That all schemes produce the warm cloud mass, but some configurations rain it out if the autoconversion rate is fast enough, leads us to suggest that it is not the Thompson scheme per se which is responsible for producing the low-level cloud mass, but rather the meteorological conditions present in which these simulations are performed. We provide in our response to Reviewer 2's comment 21 a figure (Fig. R4 in the response to Reviewer 2) figures and further discussion on the presence of this low warm cloud mass.

We have now included such comments on the presence of the warm cloud mass in our Results (P20 L10-16) section.

However, we disagree with Reviewer 2's comment that there is a general lack of interpretation. In the liquid cloud (warm cloud mass in the south of the Congo domain and the cloud in the RICO simulations) the differences can be attributed to the autoconversion, as we have already shown in our tests and describe in the discussion of our figures. In the deep convective cloud (where ice processes can occur), much of the difference between the schemes can be attributed to differences in the classification of frozen particles, which we have already discussed at length.

In addressing the Reviewer's comments on lack of interpretation, we have now performed further simulations which test the autoconversion of cloud ice to snow in the two schemes. These are presented and discussed in our revised manuscript and can be found from P20 L17 to P21 L2.

5. The section of Conclusions is weak. The authors stress that their key findings are a) "A key finding is that the simulated hydrometeor classes differ significantly between microphysics schemes" b) "Another key finding is that the difference between the hydrometeor classes simulated by each microphysics scheme varies between cases of convection." c) Another key finding is that the cloud morphological difference and the difference in the hydrometeors between different schemes is significantly larger than that due to CDNC perturbations.

As it was said above, all these findings are not new.
I would recommend to rewrite conclusions by adding more detailed analysis of results and recommendations of the ways to improve the schemes.
We present the new results that the way in which the schemes differ from each other between cases of convection is not systematic, and further that their response to aerosol also differs non-systematically.

By performing simulations within a single modelling framework we reduce much of the uncertainty in comparing results from different microphysics schemes noted by Khain 2015, such as orography, boundary layer parameterization, etc. We note that we have only performed real-data simulations in one region over one 10-day period in August 2007 and therefore would not expect our results to be exactly the same in another region or in another season, where orography, large-scale meteorology, and cloud type would be different from that present in our simulations. However, our main result is that cloud impacts from choice of microphysics scheme far exceed cloud impacts from CDNC perturbations in every case we consider, and therefore we would expect that even though the cloud development and response to CDNC would clearly be different in another regime, the uncertainty due to choice of microphysics scheme would still dominate.

We have rewritten our Conclusions, and put our results in the context of a wider body of previous work. These can be found in the revised manuscript from P21 L21 through P26 L4.

We stress that without observations it is difficult to suggest ways to improve the schemes. However, we have now performed extra tests to explain the processes which lead to the large differences in upper-level ice in the Congo simulations. This is discussed in our Results section from P20 L18 through P21 L2, along with suggestions as to how the schemes could be improved.

Further, we also find the same persistent upper-level ice in our WRF-SBM Congo simulations, which we present in the Supplement. Although the focus of this paper is not to provide a comparison of bin vs bulk schemes, we show that the differences resulting from conversion of one ice category into another is a limitation of any scheme whether bin or bulk which uses fixed ice categories. Our results support the argument that ice phase processes may be better represented if developments in microphysics schemes starts to move away from the use of fixed ice categories.