REPLY TO REVIEWER #1:

We thank the reviewer for the comments that helped us to improve the manuscript.

1. The logic (the same as applied in Seifert et al. 2010) is to derive the autoconversion and accretion enhancements for the 2-moment scheme of Seifert and Beheng, and then to use the modified 2-moment scheme in LES simulations. I feel this is a justifiable methodology (especially considering the expense of the bin scheme), but I feel the 1D kinematic model of Seifert and Stevens might not be sufficient to validate the 2-moment implementation. To me, the key difference between bin and 2-moment scheme is the representation of droplet sedimentation (mass/number weighted in the 2-moment scheme and different for every bin in the bin scheme). Thus, the surface rainfall (e.g., Fig. 4 in the current manuscript) may agree well in the 1D test, but may differ more significantly in a test where horizontal variability is included, for instance, in a 2D kinematic test. Overall, I feel the difference in the sedimentation between bulk and bin schemes deserves a closer look, not necessary in the context of the current paper, but in a more general study. I would like to see this aspect at least to be recognized in the current draft.

We agree with the reviewer that a 2D framework would be a much better test and fully agree with the statement concerning sedimentation. We have added a sentence at the beginning of section 4 reading

Although the 1D model provides a reasonable idealized framework for such a test, we would recommend to use a kinematic 2D model (e.g. Szumowski et al. 1998, Morrison and Grabowski 2007) in future studies, because the 1d framework might not be sensitive enough to differences in the treatment of sedimentation which are more relevant in a more complex flow field. Here we apply the simpler 1D model for consistency with Seifert et al. (2010).

2. The fact that differences in the cloud microphysics (i.e., rain formation in the current study) may affect cloud dynamics is obvious. However, this aspect is not even mentioned in the current

manuscript except for (relatively obscure and not discussed) references to the inversion height shown in Fig. 6. I think some discussion of the feedback from the microphysics to the cloud field dynamics (e.g., deepening of the cloud field that is an unfortunate feature of the RICO setup) should be added to the manuscript. Overall, separation of purely microphysical effects from the impact on cloud dynamics is difficult, but needs to be done to fully understand the impacts. Again, I feel just men- tioning this issue and leaving it for a future study (perhaps applying the piggybacking method that Grabowski used in his studies published in JAS in 2014 and 2015) would be sufficient. A hint of the dynamic feedback can perhaps be shown by adding the inversion height to time evolutions shown in Fig. 5.

The deepening of the cloud layer is one of the most interesting features of the RICO case and makes it especially valuable when investigating the effects of cloud microphysics on the evolution of the cloud layer. The effect of different microphysical choices or assumptions on the boundary layer dynamics has been extensively discussed by Stevens and Seifert (2008), van Zanten et al. (2011), Seifert et al. (2015) and others. Therefore we have not discussed this in detail in the current manuscript. In the revised version we follow the recommendation of the reviewer and have added the inversion height to Fig. 5 and included a few sentences in section 5.2. reading

The main feedback of the different microphysical developments on the dynamics and evolution of the boundary layer as a whole is that rain formation arrests the growth of the cloud layer as it can be seen in the time series of the inversion height in Fig. 5, i.e., the Ayala-Wang kernel leads to a much shallow cloud layer in the precipitating regime. A similar behavior for different cloud droplet number densities was shown by Stevens and Seifert (2008). For the RICO case the boundary layer deepens and supports successively deeper clouds until moisture is efficiently removed by precipitation. Eventually the precipitating regime reaches a quasi-stationary state, the subsiding radiative-convective equilibrium (Seifert et al., 2015).

Using the piggybacking methodology would be an attractive alternative to our extensive LES study. Without piggybacking the randomness of the individual LES runs makes it actually necessary to use ensembles of LES realisations, which is computationally very demanding. We agree with the reviewer that piggybacking offers an attractive method to overcome such problems. Nevertheless, we refrained from using the method because it leads to inconsistencies between the dynamics and the microphysics and the results have to be interpreted very carefully. The old fashioned brute force approach used in our study is maybe less elegant, but each simulation is physically fully consistent. Nevertheless, we fully agree that such studies as presented in our manuscript could benefit from the piggybacking approach, if it is carefully used and interpreted.

3. P. 3, paragraph starting at l. 30. The way enhancements are shown in Fig. 1 does not allow seeing the enhancement for droplets of equal (or very close) size. Can you show the enhancement for equal-size droplets for the two formulations? How important are such collisions for the acceleration of rain formation?

The enhancement factor for equal-size droplets is by definition infinite. We would refer to Fig. 4 and section 4.3 of the accompanying paper by Onishi and Seifert (2016, ACP) for a discussion of the collision frequency of similar sized droplets. We think that such collisions, e.g. selfcollection events of small raindrops, are very important especially in maritime clouds with low to moderate cloud droplet numbers and relatively high autoconversion rate. In such clouds small drizzle drops can be present in abundance, but their growth is relatively slow due to the low to moderate cloud water content (limiting accretion) and the rare collisions between similar sized drops (limiting selfcollection). As soon as some drops grow due to some selfcollection events, they also have an advantage in accretion due to the larger fall speed of a bigger drop. Such a chain of processes is what we postulate to explain the increase in accretion rate (Eq. 15), which is stronger than the enhancement of the kernel itself for the accretion process. Or in other words: The enhancement of the collision rate of similar-sized drops leads to a modification of the drop size distribution (a stronger tail) due to selfcollection which is part of the enhancement of accretion parameterized by Eq. (15).
The importance of selfcollection for the surface rain rate in maritime shallow cumulus is also discussed in the recent paper by Naumann et al. (2016) by applying a detailed diagnostics using a Lagrangian drop model (aka superdroplets).

4. P. 4, the end of section 4. I think you can explicitly say when discussing Fig. 4 that the differences are about 10-20 % max, a relatively small difference considering differences seen in cloud field simulations.

Figure 4 is not only there to show that the bulk scheme works reasonably well, but also and maybe more important to discuss the differences between the two collection kernels. It is not clear to which of the two the reviewer refers. The difference between the Ayala-Wang kernel and the Onishi kernel can actually be a factor of 2 (for moderate dissipation rates).

5. P. 7, discussion around l. 29. I think the discussion has to do with the undesirable aspect of the RICO case, namely, the deepening of the cloud field. Perhaps this should be openly stated (I think it is not obvious to someone not familiar with the RICO case). My suggestion at the end of 2 above would also help to make this obvious.

Following the recommendation of the reviewer, we have included a discussion of the deepening of the cloud layer in section 5.2. Nevertheless, we do not understand why the deepening of the cloud field should be 'undesirable'. As long as the subsidence drying is not able to compensate the moisture input from the latent heat flux the cloud layer has to grow. We could agree with the statement that the growth of the cloud layer is artifically slow in the RICO case making it much more susceptible to microphysical perturbations than a boundary layer in which local radiative cooling leads to a more rapid equilibration of the cloud layer, i.e., the deepening should be much more efficient than in the standard RICO case used here.

6. P. 8, text between l. 10 and 15. I feel more explanation is needed here. What is σ_x (mean standard deviation from the time average?). What is the lag-1 auto- correlation? How many samples are there in the 6-hour time series? This method of assessing statistical significance is different from the Student t-test statistic, correct?

Yes, the domain mean quantities are simple time series and σ_x is the standard deviation as it is explained in the text. The standard deviation of a time series is always 'the mean standard deviation from the time average'. The lag-1 autocorrelation of a discrete time series is the autocorrelation between subsequent samples of that time series. This is standard terminology in statistics and time series analysis (and easily found in most textbooks). Software packages like R, Matlab, NCL, etc. provide functions to calculate these quantities. The estimation of the effective sample size is a classic problem in statistics and the reference provided in the paper gives a more detailed discussion of this topic.

The number of independent samples depends on the quantity, because different variables have different autocorrelation time scales. For the rain rate the effective sample size in a 6-hour time series is between 3 and 10 with an average of about 6. This makes sense as a shallow convective rain event has a typical duration (or time scale) of 1 hour. For the inversion height the sample size is only 1 per 6-hour time series, because the inversion height is the result of the combined action of all boundary layer eddies (i.e. all clouds), i.e., each LES run provides only 1 independent estimate for the inversion height. Due to this averaging property the standard deviation of the inversion height is also much smaller and consequently the standard error is small although the effective sample size is only 1 per LES run. Knowing the effective sample size is a prerequisite for the Student t-test, but we decided to plot only the standard error and not to delve deeper into test for statistical significance. We would argue that even without doing statistical hypothesis testing our analysis is still more elaborate than what is usually presented when comparing different LES runs.

7. P. 10, l. 30. Here is an example of the microphysics-dynamics feedback that is important in this problem, yet it is really not discussed in the current draft.

This feedback is now mentioned several times in the revised manuscript. For a detailed discussion of the basic behavior we refer to the literature, e.g., Stevens and Seifert (2008) as well as Seifert et al. (2015).