We want to thank reviewer 1 for his/her careful reading of the manuscript, the constructive and valuable comments. This document contains point-to-point replies to each point the reviewer made. The reviewer's comments are written in normal, our replies in bold italic font. Line numbers preceded by an exclamation mark refer to the original manuscript, those without to the revised one.

The authors compare three different Langrangian Cloud Model (LCM) implementations with a focus on collection using three different collection kernels. Analytical solutions as well as previous bin model results are used as references. Additionally, sensitivity of the LCM implementations with respect to the initialization of the simulation particles (SIPs) as well as to different numerical features (resolution, time step, ...) is tested.

This results in a large amount of model runs with a great variety of possible parameter and configuration combinations which sometimes makes the manuscript difficult to read.

General comments

Each of the LCM implementations shows rather strong shortcomings:

• RMA cannot deal with realistic kernels (Long, Hall) and shows spurious oscillations.

• AIM systematically underestimates the collisional growth.

• AON always needs an ensemble of at least 50 realizations to reach a representative average result for the final drop size distribution since individual realizations deviate considerably from the average (in contrast to RMA and AIM). This severely limits the potential to be used in 2- and 3-dimensional models with a large number of grid points.

We agree with the reviewer's opinion on RMA and AIM. For AON we are not as pessimistic as the reviewer regarding its use in 2-D and 3-D models. It will not be necessary to run 50 realisations in each grid box. In such cases, the averaging will occur over grid boxes with similar atmospheric conditions. We made a similar experience at least in simulations of contrail-cirrus, where we tested the NSIP-sensitivity of the deposition/sublimation process (see section 3.1 in Unterstrasser & Sölch, 2014). We found that very few SIPs per grid sufficed to reach convergence even though the few SIPs in a single grid box could not realistically represent a smooth size distribution. Smooth size distributions can be derived only for larger volumes of air when more SIPs are taken into account. This explanation is now given in section 4.1.

It will be interesting to see how AON performs in 2D/3D-setups. This will be the next step in an upcoming study.

Additionally, sensitivities with respect to initialization of SIPs are shown to be high at least for some configurations. This problem is discussed towards the end of the manuscript where also more mature drop size distributions are used for initializations.

Within a full microphysics description including drop nucleation and condensational growth, it should be harder to control the DSD at the moment when collisions become important. This discussion

We agree and it will be interesting to see how the algorithms behave in such a setup and what extensions may be necessary to guarantee optimal SIP weights. SIP splitting and merging as discussed in Unterstrasser & Sölch, 2014 may be an option.

should be extended. Compared to spectral bin models the accuracy of all LCM implementations shown seems to be lower with at least comparable computational costs. One could conclude that LCMs are of no practical use. Nevertheless, LCMs are valuable tools. Please discuss critically advantages and disadvantages of LCMs.

In the Introduction we add a paragraph mentioning the most important advantages/disadvantages of LCMs: "Due to their specific construction, LCMs offer a variety of advantages in comparison to spectral-bin and bulk cloud models. Their representation of aerosol activation and subsequent diffusional growth follows closely fundamental equations and avoids therefore the possible perils of parameterizations (e.g. Andrejczuk et al. 2008, Hoffmann 2016). The same applies for the representation of collection or aggregation, which is based on the interaction of individual SIPs. Accordingly, LCMs approximate pure stochastic growth (e.g. Gillespie 1975), which is the correct description of collection/aggregation within a limited system of interacting particles and results in the SCE, which is used as the basis for spectral-bin and bulk models, if the system becomes infinite (e.g. Bayewitz et al. 1974). Moreover, LCMs do not apply the finite-differences method to compute microphysics. Accordingly, LCMs are not prone to numerical diffusion and dispersion, and do not suffer from the numerical broadening of a droplet spectrum, which can affect spectral-bin cloud models (Khain et al. 2000). Finally, LCMs enable new ways of analysis by the tracking of individual SIPs. They can be used to reveal the origins of droplets, as well as conditions associated with their growth (e.g. Hoffmann et al. 2015, Naumann and Seifert 2016). The largest disadvantage of LCMs, so far, might be their relative novelty due to their higher computational demand. Many aspects of this approach have not been validated adequately or can be improved. For the process of collection/aggregation, this study will offer a first rigorous evaluation of the available numerical approaches. "

It is clear, that the present study is a first step of evaluation. Our next step is the AON evaluation in higher-dimensional tests. For this reason, we do not want to speculate too much about the performance in such applications. The initialisation of SIPs differs from model to model and/or application to application. Constant weights approaches, as used in several studies recently, greatly deteriorate the collection treatment. Those studies initialise the SIPs in the beginning of the simulation as CCN. Contrary to this, the LCM by Sölch and Kärcher, 2011, initialises SIPs only during cloud formation with varying SIP weights.

Regarding the computational costs no final conclusion can be made at the present stage. At least for localised cloud objects like contrails as studied with EULAG-LCM, the computational costs of LCMs and bin models are comparable. A bin model would carry out the same amount of computations in any grid box, whereas in LCM approaches no computations are carried out in the ice-free grid boxes (which can be the majority). On the other hand, this might not necessarily result in a model speed-up due to load-imbalancing.

The quality of some figures is poor. Most of them are too small, lines are too thin and sometimes too many. Specific comments are given below.

All figures have been revised.

Specific comments

l. 154-157: What is the reason to switch from mass doubling (which is often used) to a tenfold increase as a basis for bin resolution?

There is no particular reason besides personal flavour. A simple conversion formula is given in l. !157.

l. 238: If the probabilistic version of the singleSIP-init is used, dots are not distributed uniformly!

Strictly speaking, the dots are uniformly distributed only for the deterministic version. We hope that saying "homogeneously distributed" is an acceptable expression.

Fig. 1: upper left and below: difference between red and green lines is misleading since the higher density of dots wrt the x-axis is not resolved.

In the updated figure, κ is reduced such that the dots appear less connected.

upper right: threshold radius line barely can be seen; lines for alpha-values are also misleading, can be confused with legends. Values should read N 10α .

Thank you for spotting the inconsistency with the usage of α. We improved the figure.

Last but one row: Is there a systematic difference between the symbols and the lines due to plotting issues? If not, a better initial agreement should be reached. Cp. also l. 268-270.

The reason for the systematic difference between the symbols and the lines is the following: The data of the reference solution use a finer bin grid. Now we use the same κ as for the plots with the symbols and the agreement is better. As noted in the manuscript, we use $\kappa_{plot}=4$.

Algorithm 1: What do k++ and i++ stand for? loops over k/i?

This is C++ style for "k = k +1". We replace those expressions.

Algorithm 2: l. 13: Please exchange gain and loss term due to consistency with l. 12 and eq. (22)

Thanks for this hint. Corrected.

Fig. 3: Top: It is difficult to see what happens in the left part of the spectrum (<20 μm). Bottom: It is confusing to normalize one ratio to t=0 and the other one to t=3600. Please redo the black curve with vi (t = 3600)/vi(t = 0)

The motivation behind the definitions was that both fractions are >= 1 and both curves fit better in the same plot. We changed the description of the plot around line 473 to eliminate this pitfalll.

l. 559 and Figs. 6, 9, 11, 13, 15, 17, 19, 22: The third moment λ3 should not be shown in the figures since the behaviour is very similar to λ2 (which should be stated in the text). The space saved can be used to extend some other figures in order to improve their readability.

We remove all panels that show the third moments.

Fig. 5: Use full lines with enhanced line thickness for RMA results and dotted (or dashed) lines for analytical solutions. Otherwise, all plots look identical at first glance.

We changed it in all plots showing DSDs. Now the solid lines show our results, the dotted lines show the reference solution.

l. 564-575: The discussion of the RMA Golovin results is very short and misses several aspects, e.g.: Why are the results for RedLim worse than the regular ones? What are the reasons for the relatively large differences between the two OTF versions?

We expanded the description of the results in section 3.1.

Fig. 6: Are there any lines missing? Variation of η only for κ = 60 in the left and the middle column? No κ = 60 for OTF at all? Which lines fall together and which runs are carried out at all?

We added the necessary information in the caption to avoid any confusion. The variation of η was indeed only shown for κ =60. And in the right panel only four simulations were shown. We redesigned the plot and the selection of simulation at display changed slightly.

Fig. 7: see Fig. 5: full lines for the RMA results.

Done.

l. 595: Compared to the regular version (and to bin model results) I would not call the RMA RedLim results "perfect". The same holds for the OTFs results; only OTFl is almost "perfect".

We agree and reformulated the paragraph in section 3.1.

Figs. 8, 10, and 12: see Fig. 5: full lines for the AIM results.

Done.

Fig. 14: see Fig. 5: full lines for the AON results. Results plot for disregarding self-collections is missing.

Good point. We decided at some point to leave out the third panel and forgot to change the caption.

l. 739: Is this restricted to AON results or do the other methods show similar robustness wrt to the small tail of the distribution? In reality very small drops similarily do not contribute substantially to the growth of the large mode due to their low number and small individual mass. This should be reflected in model sensitivities.

This insensitivity to the small tail is even more pronounced in the AIM algorithm. This was already mentioned in the original manuscript around line !642.

l. 746: This is in contrast at least to the Golovin RMA results. Why is it reasonable for AON? Is this due to the lower number of collision events realised because of the probability restrictions?

Yes, this is the reason. We added an explanation around line 815.

l. 766: When pcrit is smaller, less collection events can be expected (see lines 469/470). A spread in v-values leads to smaller and larger v-values. Does this mean that the largest v-values are responsible for the enhanced collection?

The sentence contains a typo. "p_{crit} is smaller" must be replaced by "p_{crit} is larger". A combination of a small v SIP and a large v SIP leads to a large p_{crit}.

l. 902ff: It should be critically mentioned that AON always needs an ensemble of at least 50 realizations to reach a representative average result for the final drop size distribution since individual realizations deviate considerably from the average (in contrast to RMA and AIM). This leads to a large effort in terms of computational resources.

See our response in the beginning of this reply. We included our thoughts in section 4.1.

Technical corrections

Thank you for spotting all those typos and inconsistencies. They have been corrected.

l. 18: ... are important processes ...

Table 1: mean mass: M/N

Fig. 1 caption: alpha should be (-2, -3, -7)

l. 256: values of N 10α ...

l. 276: However, it is ...

l. 371: rather "proposed" than "discussed"

l. 410ff.: The terms "larger SIP" and "smaller SIP" are used here with the meaning "SIP with larger/smaller drops(=higher average drop mass)". Please define whether "large SIP" indicates large drops or a large number of drops within the SIP (cp. l. 510).

We intended to use large/small SIP throughout the text in the sense you mention. We added a sentence on the terminology around line 120.

Fig. 3 caption: ... function of their initial radius ... Please add that it is an AIM simulation.

l. 434: ... of each droplet within the SIP ...

l. 435: Figure 3

l. 510: In contrast to l. 410ff. smallest SIP refers to the size of the droplets not the weighting factors;

The description is correct the way it is. In both paragraphs small SIP refers to a SIP with small droplets and the terminology is consistently applied.

Fig. 11 caption: ... black curves with triangles ... green lines for vrandom; α should be (-2, -3, -7)

l. 658: green lines

l. 792: check the meaning of "large SIP", also l. 824 "heavy SIP"

Again large SIP refers to a large single droplet mass. Heavy SIP are SIPs with large total mass $\chi = \nu \mu$.