Reply to Reviewers

July 14, 2017

We would like to thank the reviewers for their questions and comments. Before we answer them, we need to point out that we found an inconsistency in the way the collision efficiency tables were interpolated in the SDM and in the Bott method. It only affected simulations that use efficiencies from Hall (1980) for large droplets and from Davis (1972) for small droplets. This kind of collision kernel was used in Sections 5 and 6. The SDM simulations presented in Figs. 7, 8 and 9 were repeated with the problem fixed. The main difference is that now we see convergence of the "one-to-one" results to the Smoluchowski equation in the slow-coalescence case. What did not change is the fact that using larger coalescence cells can decrease the rate of conversion of cloud to rain drops due to additional collisions between rain drops. In consequence, using the Smoluchowski equation can underestimate the amount of rain produced. The problem affected only large drops, with radius greater than ca. 90 microns. Therefore the lucky droplet analysis from Sec. 6, in which droplets grow only up to 40 microns, remains valid.

Answer to the Anonymous Referee #1.

- The methodological section (2) should be expanded. From the current text, it is not possible to detect the equation of motions for the single droplets. Are the droplet tracers or inertial? Or are they just subjected to the gravitational force? A complete set of kinematic, dynamics and radius equation evolution should be given for a more general configuration, and then the system can be simplified depending on the hypothesis introduced by the authors.

We use box model simulations, which are convenient for studying coalescence. Droplet motion is not modelled, therefore we do not give their equations of motion. However, we use gravitational coalescence kernels, so droplets collide as if they settled due to gravitation. To clarify it, Section (2) was rewritten and says:

" Consider coalescence of water droplets in a well-mixed volume V. Other processes, like water condensation and evaporation, are not included. Thanks to the assumption that the volume is well-mixed, all droplets within the same well-mixed volume can collide with each other, independently of their positions (Gillespie, 1972). Therefore droplet motion does not have to be explicitly modeled and droplet coalescence can be calculated in a stochastic manner, as it is done in the master equation. Consider two randomly selected droplets i and j. Probability that they collide during timestep Δt is $P(r_i, r_j) = K(r_i, r_j)\Delta t/V$, where r_i and r_j are their radii, K is the coalescence kernel and V is volume of the box. We use gravitational coalescence kernels, so the effect of turbulence on coalescence is not studied. "

- The English level of the manuscript needs to be improved. The most frequent error is the lack of articles in front of many substantive in all the manuscript (in collisional growth in the collisional growth, growth rate of lucky droplets The growth rate of lucky droplets, just to give few examples). The origin derives from the lack of articles in Slavic languages, so the manuscript English level should be more carefully addressed in the next revision.

We made an effort to improve the text. If the manuscript is accepted, remaining errors will be fixed during the copy-editing that is included in the processing charges.

- The authors compare the DNS case by Onishi et al. (2015). It is not clear in the paper how the comparison has been done. Again, it is not clear if the super-droplets are influenced or not by (and if they move driven by) turbulent fluctuations or if the comparison is done just considering gravitational settling. The latter case would imply that the turbulent fluctuations have a weak effect on particle-particle collisions that it does not seem the case in reality.

We studied coalescence only due to gravitational settling. It is now written explicitly in Section 2:

"We use gravitational coalescence kernels, so the effect of turbulence on coalescence is not studied."

Onishi et al. (2015) performed DNS both for stagnant and turbulent air. They found that the mean autoconversion time is significantly decreased by turbulence, so turbulent fluctuations do have strong influence on collisions. However, we compare with them not the mean autoconversion time, but the relative standard deviation of autoconversion time and how it scales with the system size. Onishi et al. (2015) show that turbulence can change the relative standard deviation of autoconversion time by about 25%. While this is a significant change, it is barely visible on the logarithmic scale in Fig. 5.

-Many important references are missing regarding the methodology: Unterstrasser et al. 2016 doi:10.5194/gmd-2016-271, Li et al.

2017 doi:10.1002/2017MS000930. The latter, in particular, has many analogies and supports the results of the current manuscript.

Both papers are now cited in Section 1:

" A thorough comparison of coalescence algorithms from Lagrangian methods was done by Unterstrasser et al. (2016). It lead to the conclusion that the method of Shima et al. (2009) "yields the best results and is the only algorithm that can cope with all tested kernels". It was also found to be optimal in DNS tests (Li et al., 2017)."

Moreover, Li et al. (2017) is now cited in Section 8:

" Li et al. (2017) have shown that condensation can regulate differences between Eulerian and Lagrangian coalescence schemes. Discrepancies between these schemes that they observed in simulations with condensation and coalescence were smaller than in pure coalescence simulations. "

-It would be interesting to include the effects of condensational growth as stated in the last paragraph of the conclusion. A new Lagrangian stochastic model has been proposed by Sardina et al. 2015 doi:10.1103/PhysRevLett.115.184501. The model could be easily implemented in the super-droplet framework.

We have the option to include condensation in our model. We did not do it, because we believe that it is important to first understand the simpler problem of pure coalescence before dealing with more complex problems.

-The LES sentence in the introduction can be obscure for non-specialist researchers in the field. The implication of the current approach for LES can be fundamental. The paragraph should be expanded to explain better the concept of LES and why collisions should be accurately modelled in the absence of small turbulence scales.

We do not include turbulence in our coalescence scheme, so we think that there is no need to explain the concept of LES. We agree that the sentence could be obscure. Moreover, use of the super-droplet microphysics is not limited to LES. For these reasons we changed the sentence so that it does not mention LES anymore. Implications for cloud modeling, including LES, are discussed in Sec. 8.

-Section 6-Lucky droplets: The values of Kostinski and Shaw (2005) are estimation. The sentence: their theoretical analysis overestimates the luckiness in droplet growth is too strong, the order of magnitude of their analysis is the same of the one detected with the super-droplet method.

We changed that to:

"their theoretical analysis slightly overestimates the luckiness in droplet growth."

Technical corrections: -It is hard to distinguish the different lines in most of the plots if printed in black and white.

In addition to different colors, lines now also have different dashing.

-For the reader point of view, it is easier if the comparison with the results of Alfonso and Raga (2016) are embedded directly in figure 1 and figure 2 (as the authors already did for figure 5).

We have obtained the data from the authors of Alfonso and Raga (2016) and plotted it in the Figures 1 and 2.

-Figure 4: Is it possible to include in the plot the DNS results for a better comparison?

Onishi et al. (2015) give a DNS result for stagnant air only for one system size. We have added it to the Figure 4. The DNS result is significantly different from the SDM and the Smoluchowski equation results. Following Onishi et al. (2015), we conclude that this is due to the inaccuracy of the Hall coalescence kernel that was used in the latter two. Part of Section 4 that discusses Figure 4 now says:

" The SDM results are also compared with the results of DNS, in which air turbulence was not modelled, but hydrodynamic interactions between droplets were accounted for. We choose this kind of DNS, because it should be well described by the Hall kernel that is used in the SDM and in the Smoluchowski equation. It turns out that the Hall kernel gives too short autoconversion times. The same issue was observed by Onishi et al. (2015) (cf. Fig. 1(b) therein). "

Answer to the Anonymous Referee #2.

- The presentation is too short to fully understand what has been done and to be able to judge it. Moreover, the statements are too general. Also the motivation behind choosing the presented tests must be made clearer. In the present style the presented tests resemble a bit a random collection of tests related to a specific SD coalescence algorithm.

We have made an effort to make the statements and the presentation more specific. Regarding the motivation for different tests, we have rewritten the last paragraph to the introduction to make it more clear:

" The Shima algorithm is not based on the Smoluchowski equation, but, similarly to the master equation, on the assumption that the volume is well-mixed. The Shima algorithm introduces some simplifications that may increase the scale of fluctuations in the number of collisions, as described in Sec. 2. These simplifications are not necessary in the limiting case of a single computational particle representing a single real particle, what we call "one-to-one" simulations. Then, the Shima algorithm should be equivalent to the SSA, i.e. it should produce a realization in agreement with the master equation. To show that this is true, we compare the Shima algorithm with the master equation and the SSA in Sec. 3. We also compare it with the more fundamental DNS approach in Sec. 4. Once the "one-to-one" approach is shown to be at the same level of precision as the master equation, we use it to study some physical processes that are related to the stochastic nature of coalescence. The way the sol-gel transition time changes with system size is studied in Sec . 3 and in Sec. 6, we quantify how quickly the luckiest cloud droplets become rain drops. In addition, we use the "one-to-one" approach to validate more approximate methods. The Shima algorithm with multiplicities greater than one is studied in Sec. 4. We determine how many computational particle s are required to obtain the correct mean autoconversion time and correct fluctuations in the autoconversion time. Next, in Sec. 5, we determine how large the system has to be for the Smoluchowski equation to correctly represent the rate of rain formation. Throught the paper we observe that evolution of the system strongly depends on its size. The size of a well-mixed air parcel is estimated in Sec. 7 and some implications for cloud simulations are discussed in Sec. 8. "

In addition, we changed the titles of Secs. 3 and 4 to make it more clear what is their purpose.

General points 1. The description of your method and results is too short and often not clear enough. For a reviewer it is difficult to understand what you have done in detail and this makes it hard to thoroughly rate your work. The description in section 2 must be considerably expanded and be more precise. Also the results section should be improved. For example the motivation behind showing the comparison in section 3 is not really clear to me. As an example I go over page 3 and try to highlight what I miss there and where I suspect that other readers would get stuck:

The method was described in detail by Shima et al. (2009), so our intention was only to describe how our simulation method differs from theirs. Apparently this makes the method not clear, so we extend the description of the method as asked by the Reviewer in the following points. Regarding the motivation behind different comparisons, it is now given in the Introduction, as explained in the answer to the previous comment.

It is not always clear if you talk about real droplets or super-droplets. It should be always clear if your statement relates to the real world or the super droplet world.

It is now clarified in Sec. 1:

" We will refer to these computational particles as super-droplets (SDs). The words "droplets" and "drops" are reserved for real hydrometeors. "

You should define the coalescence probability of a SD pair.

It is now given in Sec. 2: " Probability of coalescence of two SDs i and j that belong to the same collision pair is $P_{SD}(r_i, r_j, \xi_i, \xi_j)$ =max$(\xi_i, \xi_j)P(r_i, r_j)(N_{SD}(N_{SD}-1)/2)/\lfloor N_{SD}/2 \rfloor$ (Shima et al., 2009).

"

Can collections occur between droplets in one SD?

We use gravitational coalescence kernels, so droplets in one SD all have the same terminal velocity and therefore cannot collide. We now say it explicitly in Sec. 2:

" Real droplets represented by the same SD cannot collide with each other, because they have the same sedimentation velocities. "

l.9: Do you use two types of simulations or is it only the initialisation that differs between the two types? Then I would not call it types of simulations.

We use two types of simulations. The main difference, besides the initialisation, is that in the "one-to-one" simulations the timestep is adaptive, as stated in Sec. 2:

"Timestep length is adapted at each step to ensure that none of the collision pairs has coalescence probability greater than one."

In the "constant SD" simulations the timestep is constant and multiple collisions between SDs in a single timestep are allowed, as said in Sec. 2:

" In this type of simulation, the time step length is constant $\Delta t = 1$ s. It is not adapted, as it is done in the "one-to-one" simulations, to make the simulation computationally more efficient. Using constant time step length can make the coalescence probability exceed unity. If it does, it represent multiple collisions between a pair of SDs (Shima et al., 2009). "

In general I can imagine how the constant SD initialisation works, however I am confused by your description. o l.21: do you mean large or small NSD? Or large r?

As written, we mean large N_{SD}. For large N_{SD}, Δl_r becomes small and therefore r_{max} is small.

o rmin and rmax are defined by the relation in line 18. Using the ln(r) definition of line 19, you end up with implicit equations in the two variables. I am pretty sure this is not what you wanted to say.

Yes, we end up with implicit equations for these variables. Many pairs of values of r_{min} and r_{max} could satisfy them. We find our solution numerically, what is now explained in more detail in Sec. 2:

" The first step of the initialization is finding the largest and smallest initial super-droplet radius, r_{max} and r_{min}. They are found iteratively, starting with $r_{min} = 10^{-9}$ m and $r_{max} = 10^{-3}$ m. We require that they satisfy the condition

$$n(\ln(r_e))\Delta l_r V \geq 1, \tag{1}$$

where r_e is either r_{max} or r_{min}, $n(\ln(r))$ is the initial droplet size distribution and $\Delta l_r = (\ln(r_{max}) - \ln(r_{min}))/N_{SD}$. In each iteration, if r_{min} (r_{max}) does not satisfy (1), it is increased (decreased) by 1%. "

o You initialise additional SDs to better represent the tails of the SDs. Adding those SDs to your ensemble, isnt it necessary to reduce the multiplicity of the SD drawn from the largest bin?

No, in "constant SD" the right edge of the largest bin is r_{max}, so effectively the distribution is cut at r_{max} and droplets with $r > r_{max}$, that would be present in the real system, are not accounted for. Adding additional SDs with $r > r_{max}$ fixes this problem and does not affect the number of droplets with $r \leq r_{max}$ (i.e. the multiplicity of SDs with $r \leq r_{max}$).

Is the refinement done at both sides?

No, only on the large radius side. We now write explicitly in Sec. 2:

" We do not add SDs from the small tail of the distribution, because very small droplets are of little importance for rain formation. "

In the end, the actual number of SDs should be higher than the nominal value NSD? How much higher?

Yes, it is a little higher, what is now written in Sec. 2:

" This makes the actual number of SDs higher than the prescribed value N_{SD}, typically by ca. 1%. "

Is the one-to-one initialisation the limiting case of the constant SD initialisation?

No, because using $N_{SD} = N_0$ (i.e. multiplicity = 1) in a "constant SD" simulation would result in relatively small r_{max} and large r_{min}. Then this type of initialisation would not represent well the given distribution. Therefore a different approach to initialisation is used in the "one-to-one" simulations.

Are the multiplicities in the constant SD approach integer values? How is the rounding done?

Yes they are, as explained in Shima et al. (2009). The rounding is done to the nearest integer. The error introduced is small, because multiplicities are high.

o **l.20: To be more explicit, you should write that one SD is created per bin and its radius is randomly selected.**

It is written in Sec. 2:
" Once r_{min} and r_{max} are found, radius of one SD is randomly selected within each bin of size Δl_r. "

o **l.23: Do you want to say that the probability is allowed to exceed 1. Why does this occur only here, and not in the one-to-one simulation? Is the time step longer? Why dont you reduce it then in this case?**

Yes, in the method of Shima et al. (2009) the coalescence probability is allowed to exceed 1. It is a consequence of keepeing the time step length constant. We added to Sec. 2:
" In this type of simulation, the time step length is constant $\Delta t = 1$ s. It is not adapted, as it is done in the "one-to-one" simulations, to make the simulation computationally more efficient. Using constant time step length can make the coalescence probability exceed unity. If it does, it represent multiple collisions between a pair of SDs (Shima et al., 2009). "

o **What are multiple collections?**

It means that, during a single time step, a pair of SDs collides more than once. The details are given in Shima (2009).

o **Do you use a constant time step throughout a simulation?**

In the "one-to-one" simulations it is adaptive, as said in Sec. 2:
" Time step length is adapted at each time step to ensure that none of the collision pairs has coalescence probability greater than one. "
In the "constant SD" simulations it is constant, what is now explicitly stated in Sec. (2):
" In this type of simulations, the time step length is constant $\Delta t = 1$ s. "

2. Your conclusions are too general. Your paper reads like a general comparison between the one and only SDM approach and all kinds of other algorithms. Coalescence (or similarly coagulation of aerosols and dust or aggregation of ice crystals) can be treated in many ways in particle-based approaches (see algorithms by Andrejczuk et al, Shima et al, Riechelmann et al, Solch & Karcher, Zsom & Dullemond, Kolodko & Sabelfeld, deVille et al). There is

not the one natural way to do this. Similarly, various algorithms exist for bin model approaches. Please make clearer which statements hold in general and which are probably only valid for your specific SDM.

Regarding coalescence, the main difference between these particle-based approaches (with the exception of the DeVille algorithm, which is based on the Smoluchowski equation) is in what is the outcome of a collision of super-droplets with multiplicities $\xi > 1$. Majority of our simulations were the "one-to-one" simulations, in which $\xi = 1$. In that case, it is straightforward what the result of a collision should be, so differences between these algorithms disappear. Moreover, like the master equation, the "one-to-one" simulations are only based on the assumption that the cell is well-mixed. The numerical trick of reducing the number of collision pairs ("linear sampling") does not affect the fluctuations, as we show in Sec. 3. Therefore the "one-to-one" simulations are quite similar to the SSA. They are at the same level of accuracy as the master equation: less precise than the DNS, more precise than the Smoluchowski equation. It is now cleary stated in the Introduction:

" The Shima algorithm is not based on the Smoluchowski equation, but, similarly to the master equation, on the assumption that the volume is well-mixed. The Shima algorithm introduces some simplifications that may increase the scale of fluctuations in the number of collisions, as described in Sec. 2. These simplifications are not necessary in the limiting case of a single computational particle representing a single real particle, what we call "one-to-one" simulations. Then, the Shima algorithm should be equivalent to the SSA, i.e. it should produce a realization in agreement with the master equation. To show that this is true, we compare the Shima algorithm with the master equation and the SSA in Sec. 3. We also compare it with the more fundamental DNS approach in Sec. 4. "

In simulations with $\xi > 1$, we use the Shima method, as it was found to be optimal by Unterstrasser et al. (2017) and by Li et al. (2017). These simulations are used only in Sec. 4 in order to determine their accuracy, as explained by an added paragraph in the introduction:

" In addition, we use the "one-to-one" approach to validate more approximate methods. The Shima algorithm with multiplicities greater than one is studied in Sec. 4. We determine how many computational particles are required to obtain the correct mean autoconversion time and correct fluctuations in the autoconversion time. "

A recent study by Unterstrasser et al compared three different SD algorithms for the collection process (your all-or-nothing SDM is among the tested algorithms). Each algorithm has its strengths and weaknesses. One major finding was that the performance depends strongly on how the SDs are initialised from a given size distribution. This has several implications: 1. an accurate description of your initialisation is needed (see point 1) and

Unterstrasser et al. (2017) found that the Shima method is optimal, what we now say in the introduction: " A thorough comparison of coalescence algorithms from Lagrangian methods was done by Unterstrasser et al. (2017). It lead to the conclusion that the method of Shima (2009) "yields the best results and is the only algorithm that can cope with all tested kernels". It was also found to be optimal in DNS tests Li et al. (2017). In the light of these results, we choose to use the coalescence algorithm of Shima (2009) in this work. "

We also observed that the way the initialisation is done is important. We believe that now our initialisation algorithm is described more clearly.

2.Are your simulations sensitive to initialisation details besides the number of SDs?

It is only sensitive to the number of SDs. Initial values of r_{min} and r_{max}, if reasonable, do not affect it much.

3. In my opinion you present five types of tests: 1. Comparisons with algorithms for the master equation (ME), 2. Comparisons with DNS algorithms 3. Comparisons with algorithms for the Smoluchowski equation (KCE) 4. Comparison with analytical results and 5. Stand-alone sensitivity tests of your SD algorithm. With each type of test different aspects of your SD method can be tested. Each reference model you compare with (Bott, Alfonso, Onishi, Kostinski & Shaw) differs in which physical processes are explicitly treated. You often miss to clearly specify the purpose of and the motivation behind the individual evaluation steps you take. This must be made clearer in the revised version.

We believe that it is now made clear by the following paragraph in the introduction:

" The Shima algorithm is not based on the Smoluchowski equation, but, similarly to the master equation, on the assumption that the volume is well-mixed. The Shima algorithm introduces some more simplifications that may increase the scale of fluctuations in the number of collisions, as described in Sec. 2. These additional simplifications are not necessary in the limiting case of a single computational particle representing a single real particle, what we call "one-to-one" simulations. Then, the Shima algorithm should be equivalent to the SSA, i.e. it should produce a trajectory in agreement with the master equation. To show that this is true, we compare the Shima algorithm with the master equation and the SSA in Sec. 3. We also compare it with the more fundamental DNS approach in Sec. 4. Once the "one-to-one" approach is shown to be at the same level of precision as the master equation, we use it to study some physical processes that are related to the stochastic nature of coalescence. The way the sol-gel transition time changes with system size is studied in Sec. 3 and in Sec. 6, we quantify how quickly the luckiest cloud droplets become rain drops. In addition, we use the "one-to-one" approach to validate more

approximate methods. The Shima algorithm with multiplicities greater than one is studied in Sec. 4. We determine how many computational particles are required to obtain the correct mean autoconversion time and correct fluctuations in the autoconversion time. Next, in Sec. 5, we determine how large the system has to be for the Smoluchowski equation to correctly represent the rate of rain formation. "

4. Your study is written in a style that implies that your SD algorithm does not solve the KCE. By comparing your SD results with KCE results you seemingly disclose shortcomings of the KCE description. In fact, SDMs also rely on KCE and your study is not really suited to question the physics of the various formulations. In my understanding, you compare different numerical strategies of solving the KCE.

The SD algorithm is not a method of solving the KCE. Contrary to the KCE, it does include correlations between number of droplets of different sizes. The "one-to-one" simulations are similar to the SSA, i.e. they produce a single trajectory that follows the master equation (c.f. Sec. 2). As such, they are more precise than the KCE, so they are well-suited to discolse shortcomings of the KCE. To us it is not clear how the "constant SD" simulations relate to the KCE and the master equation. They have been shown to give mean result in agreement with the KCE in large systems (Shima et al., 2009, Unterstrasser et al., 2016). Regarding fluctuations, using the "all-or-nothing" algorithm should amplify fluctuations, because it introduces unrealistic correlations between number of droplets of different sizes. In Sec. 4 we quantify how much the fluctuations amplitude increases.

The probabilistic nature of your SDM is inherent to your specific algorithm. The SDMs of Riechelmann or Andrejczuk are not probabilistic. On the other hand, I agree that the probabilistic component of your SDM looks like it attempts to mimic the processes in nature. Nevertheless, the two should not be mistaken. In the superdropelt world of your SDM, the probabilistic approach comes from the fact, that each superdropelt represents a large number of real droplets and you may or may not perform a superdroplet collection. This not per se related to the stochastic nature of the real world processes that accounted for in the master equation.

The probabilistic nature of our SDM has the same source as the probabilistic nature of the master equation, i.e. the fact that a collision between a pair of SDs happens with some probability, according to the assumption that the volume is well-mixed. To our knowledge, this Monte Carlo approach to collisions is used in all SDMs, including the ones of Riechelmann and Andrejczuk. If multiplicities are equal to one ("one-to-one" simulations), the SDM is as much related to the real world process as the master equation. If multiplicities are greater than one,

various SDM algorithms start to differ. In the Shima algorithm that we use, the scale of fluctuations is increased, because the number of collision trials is lower than it would be in reality. We try to quantify how much it is increased. In the Riechelmann and Andrejczuk SDMs, the fluctuations are lower than in the one of Shima. Unfortunately, they do not give mean results as good as the Shima algorithm (Unterstrasser et al., 2016).

Major points: 5. In my understanding there two mathematical descriptions of the coalescence process: The Smoluchowski equation and the master equation where only the latter accounts for correlations. You never state it explicitly but I suppose that your SD method is based on solving KCE. Many spectral bin methods exist for the KCE and in the recent past SD methods for KCE were introduced. Hence, demonstrating agreement between your KCE solver and the ME solver by Alfonso is only reasonable for a physical problem where correlations are negligible. If you want to show the numerical consistency of your approach, then a comparison with other well-established KCE solvers would be more straightforward. Could you give reasons why you compare results of your KCE solver and an ME solver? Or do you want to claim that SDMs implicitly account for correlations?

Our SD method is not based on solving KCE. The "one-to-one" simulations are at the level of precision of the master equation, what is now explicitly written in the introduction:

" The Shima algorithm is not based on the Smoluchowski equation, but, similarly to the master equation, on the assumption that the volume is well-mixed. The Shima algorithm introduces some more simplifications that may increase the scale of fluctuations in the number of collisions, as described in Sec. 2. These additional simplifications are not necessary in the limiting case of a single computational particle representing a single real particle, what we call "one-to-one" simulations. Then, the Shima algorithm should be equivalent to the SSA, i.e. it should produce a trajectory in agreement with the master equation. To show that this is true, we compare the Shima algorithm with the master equation and the SSA in Sec. 3. "

We compare "one-to-one" simulations with the master equation to validate the claim that they are at the same level of precision, i.e. that "one-to-one" method accounts for correlations. In the problem of Alfonso, correlations are very important and, as shown in Alfonso and Raga (2017), the KCE does not solve it well.

6. I am not exactly sure what you intended to show in section 3. Did you want to demonstrate the suitability of your linear sampling of droplet pairs? Then you should compare results of your algorithm, once with linear sampling and once considering all possible combination. For a simulation with 30 droplets (as done in the comparison

with Alfonso) this should be feasible. How do the curves in Fig.3 and 4 change, if you evaluate all possible combinations instead of linear sampling (at least for N0=10 and 100)?

We intended to show that "one-to-one" simulations agree with the master equation. Linear sampling is an optimization technique that we expected might be responsible for some differences between the master equation and "one-to-one" simulations. To make the comparison more detailed, now in Section 3 we compare "one-to-one" simulations, with and without linear sampling, with the master equation. We find that linear sampling does not affect mean number of collisions, nor the fluctuations in the number of collisions. Figs. 3 and 4 (up to N0=100) do not change if linear sampling is not used.

The second paragraph of Section 3 has been rewritten to explain these new results of simulations without linear sampling.

7. Is it all reasonable trying to sample a continuous exponential distribution with 10 SDs? Are the total mass and number of the SDs (averaged over all ensembles) equal to the prescribed values? What about the higher moments, do they match the values of the analytical distribution? This is an important aspect as much of your evaluated variability may come from the initialisation and not so much from stochasticity of the SDM.

Averaged over the ensamble, up to the 4-th moment of the distribution is in agreement with the prescribed one. It is true that the initial distributions can be very different between realisations and it may be the cause of large variability. For this reason we removed the $N_0 = 10$ case from Fig. 3.

8. Even though Fig.4 shows simulations with an initialisation analogous to Onishi, the presented test is a stand-alone test (following the above categorisation). Hence, further tests with other kernels or other initial size distributions can be made (as no comparison with Onishi is required) in order to corroborate your findings about how many SDs are necessary to reach convergence.

Now in Fig. 4 we compare our results with Onishi's results. Our expectation is that for other kernels relevant for cloud physics the results would not be much different. Nevertheless, we agree that tests with other kerneles, and with other initial conditions, would be useful. Such tests could easily fill a whole new paper. Our result can be considered as a guideline for users, who should do convergence tests for the specific kernels they use.

9. In Fig.4 and following figures, does the Smoluchoswki line depend on the parameter choices of the underlying Bott algorithm (time step or number of bins)? Bott is known to be diffusive; this may explain the faster generation of large droplets. Anyway I would

not call it the Smoluchowski line, it is the Bott line. The line may look differently for other bin KCE algorithms. Moreover, your lines are also Smoluchowski lines (see general point 4).

We have done convergence tests of the Bott algorithm. It is now explained in Sec. 4:

" In the Bott algorithm, we used $\Delta t = 1$ s and mass bin spacing $m_{i+1} = 2^{1/10} m_i$. The same parameters were used in each simulation presented in this manuscript. Convergence tests were done for each case. "

We agree that the Bott algorithm produces rain too soon most probably due to numerical diffusion. We now write in Sec. 4:

" The "one-to-one" results converge with increasing system volume (i.e. increasing N_0) to a value higher than the Smoluchowski result. The difference is probably caused by the numerical diffusion of the Bott algorithm. "

Labels on figures are one of: DNS, master equation, SSA, SDM ("one-to-one" or "constant SD"), Smoluchowski equation. In our view these are different appproches to solving droplet coalescence, not different numerical methods for solving some equation. The numerical methods used, i.e. the Bott algorithm for the Smoluchowski equation and the Alfonso algorithm for the master equation, are explained in text.

Minor points: 10. p.3, l21: Is the sol-gel transition an important aspect in cloud physics? You note that mass conservation is not guaranteed for some kernels. Are those kernels relevant in cloud physics? In super-droplet approaches mass conservation should be guaranteed by construction.

Mass is not conserved for the multiplicative kernel, which is not relevant in cloud physics. Nevertheless, the paper Alfonso and Raga (ACP, 2017) is a detailed study of the sol-gel transition in a small cloud volume. We decided that it will be interesting to extend their results to more realistic cases.

11. I do not understand the statement the number of collision pairs is reduced in p.4 l.4.

It was supposed to mean that linear sampling is used, i.e. less collision pairs are considered than in an exact description. Now in Sec. 2 we define the meaning of linear sampling, which is later used in the paper:

" The second simplification, that we will refer to as linear sampling, is that instead of considering all $N_{SD}(N_{SD}1)/2$ collision pairs, only $[N_{SD}/2]$ non-overlapping pairs are randomly selected. "

12. The paragraph starting p.4 l.11 should be moved to section 2.

In the paragraph it was shown that in the "one-to-one" method with linear sampling, the probability of collision between any two real droplets is the same

as in simulations without linear sampling. This does not ensure that fluctuations in the number of collisions are also correctly represented. Therefore we removed the paragraph. In its place we added to Figs. 1 and 2 the results of simulations without linear sampling, i.e. with all collision pairs considered. Their agreement with the linear sampling simulations implies that linear sampling does not affect the scale of fluctuations. Proving that was the point of the removed paragraph.

13. For Figs.1 & 2 it would be ideal to obtain the data from Alfonso and include it in the plots.

We did that.

14. P.2 l.13: Tanaka and Nakazawa present solutions for kernels other than the constant kernel. See also Table 2 in Alfonso, 2015.

We now cite Tanaka and Nakazawa in the introduction:

" The master equation was analytically solved only for monodisperse initial conditions with simple coalescence kernels (Bayewitz et al., 1974; Tanaka and Nakazawa, 1993). "

Table 2 in Alfonso (2013) gives solutions of the Smoluchowski equation, not the master equation.

15. P.5 l.5. Isnt this statement trivial? Probably any algorithm for KCE is faster than solving the master equation. A comparison among various KCE solvers would be fairer.

As explained previously, SDM is similar to the SSA and not to KCE solvers. For this reason we compare with the SSA and a solver of the master equation.

16. Figs. 2&3 show relative standard deviation of the largest droplet mass. Wouldnt it be interesting to know how large the largest droplet mass is? How many collections have occurred to form the largest droplet?

Relative standard deviation of the largest droplet mass is interesting as a measure of the sol-gel transition. We do not see a reason to show the mass of it or the number of collisions that lead to it.

17. P.7 l.10-12: I am confused. DNS simulations compute droplet trajectories and directly evaluate if there are collisions between droplets. Why do you need a coalescence kernel in this numerical approach?

It was our error. The DNS was done not for different coalescence kernels, but for different turbulence strength. We changed that sentence to:

" Small discrepancies are probably caused by the fact that the DNS included turbulence of various strength for different N_0. "

18. Fig.8: Do you also use the 20um cut-off in the Bott simulation? It is not explicitly stated in the text.

Yes, we do use it in the Bott simulation. It is now said explicitly:
" In addition, we cut the distribution to 0 at $r = 20$ m. This cutoff is used in SDM modelling as well as when solving the Smoluchowski equation. "

Technical points: Figure caption should include the information t=2500s.

We add it to the caption of Fig. 1.

P.5 l.9: droplet, not dropeltS.

Fixed.

Stochastic coalescence in Lagrangian cloud microphysics

Piotr Dziekan and Hanna Pawlowska

Institute of Geophysics, Faculty of Physics, University of Warsaw, Poland

Correspondence to: P. Dziekan (pdziekan@igf.fuw.edu.pl)

Abstract. Stochasticity ~~in~~ of the collisional growth of cloud droplets is studied in a box model using the super-droplet method (SDM). The SDM is compared with direct numerical simulations and the master equation. ~~We use the SDM~~ It is argued that SDM simulations in which one super-droplet represents one real droplet are at the same level of precision as the master equation. Such simulations are used to study fluctuations in ~~autoconversion time and~~ the autoconversion time, the sol-gel transition ~~. We determine how many computational droplets are necessary to correctly model expected number and standard deviation of autoconversion time. Also,~~ and the growth rate of lucky droplets~~is determined and~~ , which is compared with a theoretical prediction. Size of the coalescence cell is found to strongly affect system behavior. In small cells, correlations in droplet sizes and droplet depletion affect evolution of the system and slow down rain formation. In large cells, ~~unrealistic~~ collisions between rain drops ~~, caused by the assumption that the cell is well-mixed, become important~~are more frequent and can also slow down rain formation. The increase in the rate of collision between rain drops may be an artefact caused by assuming a too large well-mixed volume. Highest ratio of rain water to cloud water is found in cells of intermediate sizes. Next, we use these precise simulations to determine validity of more approximate methods: the Smoluchowski equation and the SDM with mulitplicities greater than 1. In the latter, we determine how many computational droplets are necessary to correctly model the expected number and the standard deviation of autoconversion time. Maximal size of a volume that is turbulently well-mixed with respect to coalescence is estimated at $V_{mix} = 1.5 \cdot 10^{-2}$ cm^3. The Smoluchowski equation is not valid in such small volumes. It is argued that larger ~~cells~~ volumes can be considered approximately well-mixed, but ~~only through~~ such approximation needs to be supported by a comparison with fine-grid simulations ~~. In addition, validity of the Smoluchowski equation is tested. Discrepancy between the SDM and the Smoluchowski equation is observed if droplets are initially relatively small. This implies that cloud models that use the Smoluchowski equation might produce rain too soon~~that resolve droplet motion.

1 Introduction

Coalescence of hydrometeors is commonly modeled using the Smoluchowski equation (Smoluchowski, 1916), often also called the stochastic coalescence equation. It is a mean-field equation that can be derived from the more fundamental stochastic description by neglecting correlations in the number of droplets of different sizes (Gillespie, 1972; Bayewitz et al., 1974). These correlations are especially important in small volumes and neglecting them can lead to unphysical behavior. For example, when

a single drop contains majority of water in a coalescence cell (gelation), the Smoluchowski equation does not conserve mass for some coalescence kernels (Leyvraz, 2003).

Another limitation of the Smoluchowski equation is that it describes evolution only of the expected number of droplets of given size. It does not contain information about fluctuations around this number, which are suspected to be crucial for precipitation onset (Telford, 1955; Scott, 1967; Marcus, 1968). Rate of collisions between droplets depends on their sizes. Small droplets rarely collide with each other, because they are repelled by disturbance flow induced by their settling. Once a droplet reaches a threshold size, it becomes more efficient at collecting smaller droplets. The mean time for a droplet to reach the threshold size is long, but some lucky droplets could reach it much sooner through a series of unlikely collisions. Then they grow quickly, resulting in a sooner onset of precipitation. This effect cannot be described using the Smoluchowski equation. The Smoluchowski equation can be written for the discrete number of droplets of given size, but often droplet concentration is used instead. This adds an additional assumption that the coalescence volume is large, somewhat in agreement with neglection of fluctuations and correlations in droplet numbers (Gillespie, 1972).

A number of methods alternative to the Smoluchowski equation exist. They are capable of solving stochastic coalescence, but have some shortcomings that make their use in large-scale cloud simulations impossible. The most accurate one is the direct numerical simulation (DNS). In it, trajectories of droplets are modeled explicitly and collisions occur when they come in contact. The downside of DNS is that it is computationally extremely demanding. Running large ensemble of simulations from which statistics could be obtained would take prohibitively long time. An alternative approach is to use a master equation (Gillespie, 1972). It describes temporal evolution of probability of observing a given number of particles of a given size. Collisions are allowed between all particles in some coalescence volume and are assumed to be Markovian, i.e. they only depend on the instantaneous state of the system and not on its history. This can only be justified if the volume is well-mixed, i.e. if droplets are randomly redistributed within the volume ~~between collisions~~after each collision. It is worth to note that DNS does not require such assumptions, so it reproduces correlations between positions and sizes of droplets. The master equation was analytically solved only for ~~a constant coalescence kernel (Bayewitz et al., 1974)~~monodisperse initial conditions with simple coalescence kernels (Bayewitz et al., 1974; Tanaka and Nakazawa, 1993). A more general form of the Bayewitz equation is given in ~~Wang et al. (2006)~~Wang et al. (2006), but cannot be solved for any realistic coalescence kernel. Solving the master equation numerically ~~also proves~~ is extremely difficult due to huge phase space to be considered. Recently, ~~Alfonso (2015)~~ Alfonso (2015) developed a method to solve the master equation numerically, but was only able to apply the method to a system of up to 40 droplets ~~(Alfonso and Raga, 2016)~~(Alfonso and Raga, 2017). Alternatively, the stochastic simulation algorithm (SSA) (Gillespie, 1975; Seeβelberg et al., 1996) can be used to model single trajectory obeying the master equation, but obtaining large enough statistics would require very long computations.

~~Recently, several Lagrangian schemes~~ Several Lagrangian methods have been developed to model cloud microphysics ~~in Large Eddy Simulations~~ (Andrejczuk et al., 2008; Shima et al., 2009; Riechelmann et al., 2012). Their common point is that they explicitly model microphysical processes on a small population of computational particles, each representing a large number of real particles. ~~In this paper, we use box model simulations to investigate if the super-droplet method (Shima et al., 2009) is capable of exactly representing the stochastic nature of coalescence.~~

In We will refer to these computational particles as super-droplets (SDs). The words "droplets" and "drops" are reserved for real hydrometeors. A thorough comparison of coalescence algorithms from Lagrangian methods was done by Unterstrasser et al. (2016). It lead to the conclusion that the method of Shima et al. (2009) "yields the best results and is the only algorithm that can cope with all tested kernels". It was also found to be optimal in DNS tests (Li et al., 2017). In the light of these results, we choose to use the coalescence algorithm of Shima et al. (2009) in this work. The Shima et al. (2009) algorithm is not based on the Smoluchowski equation, but, similarly to the master equation, on the assumption that the volume is well-mixed. The algorithm introduces some simplifications that may increase the scale of fluctuations in the number of collisions, as described in Sec. 2. These simplifications are not necessary in the limiting case of a single computational particle representing a single real particle, what we call "one-to-one" simulations. Then, the Shima et al. (2009) algorithm should be equivalent to the SSA, i.e. it should produce a single realization in agreement with the master equation. To show that this is true, we compare the Shima et al. (2009) algorithm with the master equation and the SSA in Sec. 3we compare super-droplet method results with master equation results for a system undergoing . We also compare it with the more fundamental DNS approach in Sec. 4. Once the "one-to-one" approach is shown to be at the same level of precision as the master equation, we use it to study some physical processes that are related to the stochastic nature of coalescence. The way the sol-gel transition . Next, we validate SDM against direct numerical simulations of conversion of cloud droplets to rain drops(time changes with system size is studied in Sec. 3 and in Sec. 6, we quantify how quickly the luckiest cloud droplets become rain drops. In addition, we use the "one-to-one" approach to validate more approximate methods. The Shima et al. (2009) algorithm with multiplicities greater than 1 is studied in Sec. 4). We also . We determine how many computational particles are needed in SDM required to obtain the correct mean behavior autoconversion time and correct fluctuations . In in the autoconversion time. Next, in Sec. 5we look for the minimal system size to which Smoluchowski equation can be applied. We use SDM in Sec. 6 to quantify how quickly the luckiest cloud droplets become rain drops and we compare the results with theoretical predictions. Finally, in Sec. 7, , we determine how large the system has to be for the Smoluchowski equation to correctly represent the rate of rain formation. Throught the paper we observe that evolution of the system strongly depends on its size. The size of a well-mixed air parcel is estimated in Sec. 7 and some implications for cloud simulations are discussed in Sec. 8.

2 The super-droplet method

Consider N_{SD} computational particles, called super-droplets (SDs), coalescence of water droplets in a well-mixed volume . Each SD is characterized by two parameters: radius r and multiplicity ξ. Multiplicity is the number of real droplets that this SD represents. Coalescence is V. Other processes, like water condensation and evaporation, are not included. Thanks to the assumption that the volume is well-mixed, all droplets within the volume can collide with each other, independently of their positions (Gillespie, 1972). Therefore droplet motion does not have to be explicitly modeled and droplet coalescence can be calculated in a stochastic manner, as it is done in the master equation. Consider two randomly selected droplets i and j. Probability that they will collide during timestep Δt is $P_{ij} = K(r_i, r_j)\Delta t/V$, where $P(r_i, r_j) = K(r_i, r_j)\Delta t/V$, where r_i

and r_j are their radii, K is the coalescence kernel and V is the volume of the box. ~~Two assumptions~~ We use gravitational coalescence kernels, so the effect of turbulence on coalescence is not studied.

At the heart of the super-droplet method is the idea that all droplets with same properties within a well-mixed volume can be represented by a single computational entity, called the super-droplet (SD). As we are interested only in droplet coalescence, it is sufficient if SDs are characterized by two parameters: radius r and multiplicity ξ, which is the number of real droplets that a SD represents. In the algorithm of Shima et al. (2009), two simplifications are made that ~~affect~~ may affect the amplitude of fluctuations in the number of collisions. The first ~~assumption~~ simplification is that SDs collide in an ~~"~~"all-or-nothing~~"~~" manner. If a collision happens, ~~all real droplets~~ each real droplet represented by the SD with lower multiplicity ~~collide~~collides with a single droplet represened by the SD with higher multiplicity. Intuitively, one would expect that this should lead to larger fluctuations than in the real system, because the number of collision trials is artificially reduced. The second simplification, that we will refer to as "linear sampling", is that instead of considering all $N_{SD}(N_{SD}-1)/2$ collision pairs, only ~~$\lceil N_{SD}/2 \rceil$~~ $\lfloor N_{SD}/2 \rfloor$ non-overlapping pairs are randomly selected. ~~The notation $\lceil x \rceil$ stand~~ N_{SD} is the number of SDs in the coalescence volume and $\lfloor x \rfloor$ stands for the largest integer equal to, or smaller than x. To keep the expected number of collisions equal to the real one, coalescence probabilities are scaled up~~be the factor $(N_{SD}(N_{SD}-1)/2)/\lfloor N_{SD}/2 \rfloor$. Intuitively, one would expect that these assumptions lead to much larger fluctuations than in the real system, because number of collision trials is artificially reduced~~. Probability of coalescence of two SDs i and j that belong to the same collision pair is $P_{SD}(r_i,r_j,\xi_i,\xi_j) = \max(\xi_i,\xi_j)P(r_i,r_j)(N_{SD}(N_{SD}-1)/2)/\lfloor N_{SD}/2 \rfloor$ (Shima et al., 2009). Real droplets represented by the same SD cannot collide with each other, because they have the same sedimentation velocities.

~~We will use~~

We will perform two types of simulations. In ~~"~~the "one-to-one~~"~~" simulations, all ~~super-droplets~~ SDs have multiplicity $\xi = 1$. ~~They~~ That way the "all-or-nothing" simplification is removed. SDs are initialized by randomly drawing radii from the assumed initial distribution. ~~Droplet collisions cause~~ Coalescence causes one of the SDs to be discarded. ~~Timestep length is chosen so~~ Time step length is adapted at each time step to ensure that none of the collision pairs has coalescence probability greater than ~~one~~1. This approach is similar to the Direct Simulation Monte Carlo method used in diluted gas dynamics (Bird, 1994). In Sec. 3 we show that the "one-to-one" method is in agreement with the master equation.

The second type of ~~simulation~~simulations, in which number of ~~super-droplets~~ SDs is constant, is closer to the ~~original idea of Shima et al. (2009)~~ideas of Shima et al. (2009). We will refer to it as ~~"constant SD"~~ the "constant SD" simulations. In ~~that~~ this type of simulations, the number of ~~super-droplets~~ SDs is prescribed, but they have different multiplicities. To avoid large differences in ~~initial droplet~~ the initial droplet size distribution between realizations, SD sizes are not completely randomly drawn from the assumed distribution as in the ~~"~~"one-to-one~~"~~" simulations. Instead, the assumed distribution is divided into N_{SD} bins and ~~size~~ the radius of a single SD is randomly selected within each bin. The first step of the initialization is finding the largest and smallest initial ~~droplet~~ super-droplet radius, r_{max} and r_{min}. They are ~~defined by the relation $n(\ln(r_e))\Delta\ln(r)V = 1$,~~ found iteratively, starting with $r_{min} = 10^{-9}$ m and $r_{max} = 10^{-3}$ m. We require that they satisfy the condition

$$n(\ln(r_e))\Delta l_r V \geq 1, \tag{1}$$

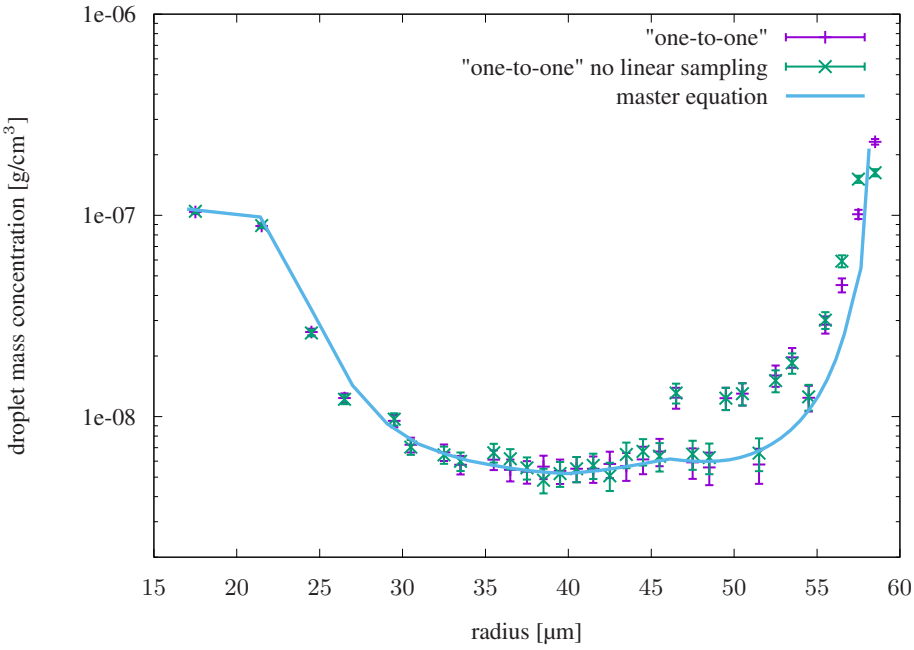
where r_e is either r_{max} or r_{min}, $n(\ln(r))$ is the initial droplet ~~sizes distribution and $\Delta \ln(r) = (\ln(r_{max}) - \ln(r_{min}))/N_{SD}$. Then, within each bin of size $\Delta \ln(r)$~~size distribution and $\Delta l_r = (\ln(r_{max}) - \ln(r_{min}))/N_{SD}$. In each iteration, if r_{min} (r_{max}) does not satisfy (1), it is increased (decreased) by 1%. Once r_{min} and r_{max} are found, radius of one SD is randomly selected ~~and its~~ within each bin of size Δl_r. Its multiplicity is given by ~~$n(\ln(r))\Delta \ln(r) V \xi = n(\ln(r))\Delta l_r V$~~. This procedure does not represent well the tails of the distribution, especially for large N_{SD}. Since the large tail is important for coalescence, we draw additional ~~$\int_{\ln(r_{max})}^{\infty} n(\ln(r))\,\mathrm{d}\ln(r) \left\lfloor \int_{\ln(r_{max})}^{\infty} n(\ln(r))\,\mathrm{d}\ln(r) + 0.5 \right\rfloor$~~ super-droplets with $\xi = 1$ from the distribution for radii greater than r_{max}. This makes the actual number of SDs higher than the prescribed value N_{SD}, typically by ca. 1%. We do not add SDs from the small tail of the distribution, because very small droplets are of little importance for rain formation. In this type of ~~simulations, coalescence probabilities can~~ simulation, the time step length is constant $\Delta t = 1$ s. It is not adapted, as it is done in the "one-to-one" simulations, to make the simulation computationally more efficient. Using constant time step length can make the coalescence probability exceed unity. If ~~they do, they~~ it does, it represent multiple collisions between ~~the~~ a pair of SDs (Shima et al., 2009). If multiplicity of a SD drops to zero, it is used to split the SD with largest ξ in the system into two. Super-droplets are discarded after collision only if all other SDs have $\xi = 1$.

We use an implementation of the SDM from the libcloudph++ library (Arabas et al., 2015). It is an open-source project available at https://github.com/igfuw/libcloudphxx.

3 ~~The sol-gel transition~~Comparison of the "one-to-one" SDM with the master equation

In a system of aggregating particles, the sol-gel transition (gelation) occurs when most of the total mass is located in a single agglomerate (Leyvraz, 2003). For some forms of the coalescence kernel, the Smoluchowski equation is known not to conserve mass after the transition. ~~Alfonso and Raga (2016) presented exact~~ Alfonso and Raga (2017) present numerical solutions of the master equation for a small cloud volume undergoing the sol-gel transition, for which the Smoluchowski equation is not valid. We perform simulations for the same setup ~~as in Alfonso and Raga (2016)~~ to test if ~~SDM can reproduce these exact results~~the "one-to-one" simulations are in agreement with the master equation approach. Consider a 1 cm^3 volume containing 20 droplets with the radius of 17 μm and 10 droplets of radius 21.4 μm. Gravitational collision kernel is used with collision efficiencies from ~~Hall (1980)~~Hall (1980). Collision efficiencies are bilinearly interpolated in the radius - ratio of radii space. Droplet terminal velocities are calculated using the formula from ~~Beard (1976)~~Beard (1976).

Figure 1 shows the average mass distribution obtained using the "one-to-one" simulations with and without linear sampling of collision pairs. In the ~~"~~latter case, all $N_{SD}(N_{SD}-1)/2$ collision pairs were considered and a constant time step $\Delta t = 0.1$ s was used. Both appproaches give the same results, what shows that the linear sampling technique does not affect the average number of collisions. In addition, the "one-to-one~~" simulations , every real droplet is represented by a single computational droplet. That way the "all-or-nothing" simplification is removed, but the number of collision pairs is reduced. Figure 1 shows the mass distribution between droplet sizes after $t = 2500$ s averaged over an ensemble of $\Omega = 10^5$ realizations. It compares well with results presented in Fig. 8 in Alfonso and Raga (2016). This implies that the "~~" simulations are compared with the master equation approach. Both approaches are generally in agreement, with some differences at the large end of the distribution.

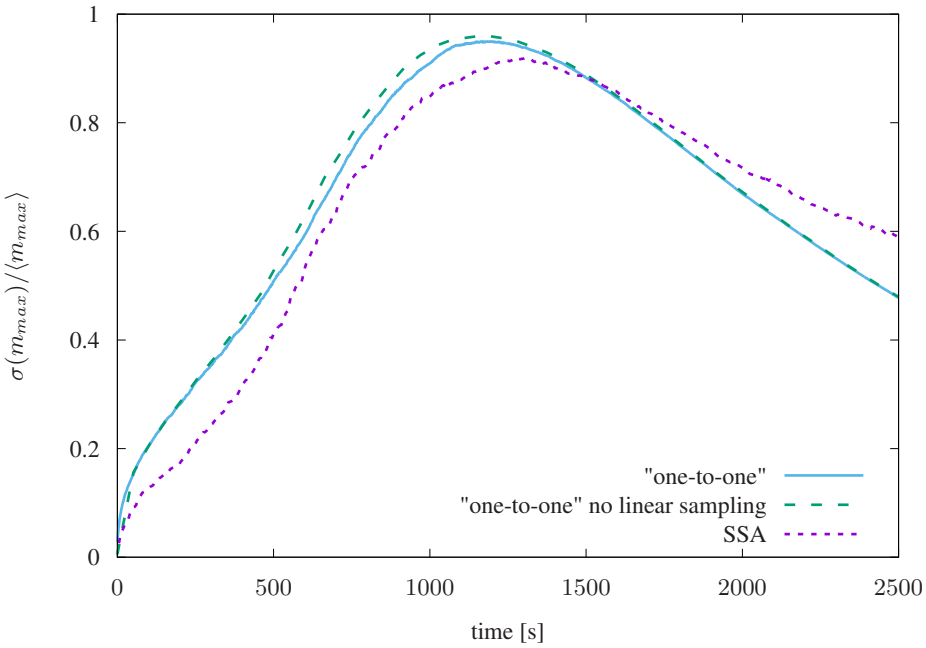


~~Mass of droplets per size bin averaged over 10^5 simulations. Bins are 1 μm wide. Shown for comparison with Fig. 8 in Alfonso and Raga (2016).~~

Figure 1. Mass of droplets per size bin at $t = 2500$ s. Bins are 1 μm wide. Points depict an averaged result of $\Omega = 10^4$ "one-to-one" simulations with and without linear sampling of collision pairs. Error bars show a 95% confidence interval. Line depicts a numerical solution of the master equation (see Fig. 8 in Alfonso and Raga (2017), data courtesy of L. Alfonso).

These differences may be caused by the way how the coalescence efficiency tables are interpolated. Another possible source of discrepancies is the numerical diffusion present in the finite-differences method of Alfonso (2015). To test if the "one-to-one~~" SDM simulation gives correct average result of coalescence, accounting for correlations between number of droplets of each size. To check if it~~" method also gives correct fluctuations in the number of collisions, relative standard deviation of mass of the largest droplet $\sigma(m_{max})/\langle m_{max}\rangle$ is plotted in Fig. 2. ~~The same plot obtained using master equation is shown in Fig. 7 of Alfonso and Raga (2016). Again, they compare very well, signifying that fluctuations are unaffected by the reduction in the number of collision pairs considered.~~ "One-to-one" simulations, with and without linear sampling, are compared with SSA simulations. As in Fig. 1, we do not observe any negative effect of using the linear sampling technique and the "one-to-one" simulations compare relatively well with the SSA. Possible sources if discrepancies are the same as in Fig. 1. Judging from Figs. 1 and 2, we conclude that the "one-to-one" approach is in agreement with the master equation approach. It accounts for the correlations in the number of droplets per size-bin and as such is more fundamental than the Smoluchowski equation approach.

~~In a "one-to-one" simulation, probability $P_o dt$ that two randomly selected droplets coalesce in a short time step dt is $P_o dt = P_{pair} \eta P_{col} dt$, where P_{pair} is the probability of these two droplets belonging to the same randomly selected pair,~~



~~Relative standard deviation of mass of the largest droplet from an ensemble of 10^5 simulations. Shown for comparison with Fig. 7 in Alfonso and Raga (2016).~~

Figure 2. Relative standard deviation of mass of the largest droplet in the system. Details of the SDM simulations are given in the caption of Fig. 1. Size of the ensamble of SSA simulations is $\Omega = 10^3$. The SSA results are taken from Fig.7 in Alfonso and Raga (2017) (data courtesy of L. Alfonso).

~~$\eta = \frac{N_{SD}(N_{SD}-1)}{2} / [\frac{N_S D}{2}]$ is the scaling-up of probability (Shima et al., 2009) and $P_{col}dt$ is the probability of coalescence if all pairs were considered. To calculate P_{pair}, we first consider even values of N_{SD}. Consider a random permutation of droplet indices. Probability that the first droplet from the pair is at an odd position in the permutation and the second is at the next position to the right is $\frac{1}{2}\frac{1}{N_{SD}-1}$. Probability that the first is at an even position and the second is to the left of it is the same . Summing these two we get $P^{even}_{pair}(N_{SD}) = 1/(N_{SD}-1)$. If N_{SD} is odd, the probability is $P^{odd}_{pair} = P^{even}_{pair}(N_{SD}-1)\frac{N_{SD}-2}{N_{SD}}$. We can write an expression for both odd and even cases $P_{pair} = 1/(N_{SD}-1+2*(N_{SD}/2-[N_{SD}/2]))$. It is readily obtained that $P_o = P_{col}$, i.e. that the probability of collision between any pair of real droplets is conserved in the "~~

The "one-to-one~~" simulations.~~

~~The super-droplet method~~ " SDM with linear sampling is computationally more efficient than solving the master equation directly, or using the SSA. It also puts no constraints on the initial distribution of droplets. Therefore we can use SDM to predict gelation times for larger systems and more realistic initial conditions. We use an initial droplet distribution that is exponential in mass $n(m) = \frac{n_0}{\overline{m}} exp(-m/\overline{m})$, where $n(m)dm$ is the number of droplets in mass range $(m, m+dm)$ in unit volume, $n_0 = 142$ cm^{-3} and $\overline{m}$ is the mass of a ~~droplets~~ droplet with radius $\overline{r} = 15$ μm. This is the same distribution as in ~~Onishi et al. (2015)~~Onishi et al. (2015). The total initial number of droplets in the system is $N_0 = n_0 V$. Results of the ~~"~~"one-

to-one" " simulations for N_0 up to ~~10^6~~ 10^5 are shown in Fig. 3. For $N_0 \geq 10^2$, the relative standard deviation of mass of the largest droplet, which quantifies amplitude of fluctuations, decreases with increasing system size. This can be understood if we look at a larger cell as an ensemble of ten smaller cells. Comparing between independent realizations, variability in the size of the single, largest droplet will be smaller if this droplet is selected from ten cells in each realization than if it was selected from only a single cell per realization. Interestingly, for $N_0 = 10^5$ an inflection point appears around $t = 500$ s. It is not seen in smaller cells. This indicates that some new source of variability is introduced. We believe that it is associated with collisions between large rain drops. We will come back to this in Sec. 5.

~~Relative standard deviation of mass of the largest droplet for different system sizes. Obtained from ensembles of 10^4 simulations.~~

The sol-gel transition time coincides with the time at which $\sigma(m_{max})/\langle m_{max}\rangle$ reaches maximum ~~(Leyvraz, 2003; Alfonso and Raga, 20~~ Intuitively, we would expect the time for most of the mass to accumulate in a single agglomerate to increase with increasing system size. This turns out to be true for systems with $N_0 > 10^3$. For system sizes $10^2 < N_0 < 10^3$ gelation time is approximately the same, around 300 s. ~~Behavior of an extremely small system with only 10 droplets is much different. Maximum relative fluctuations are smaller and gelation time is longer than in a ten times larger system. Also, the maximum of $\sigma(m_{max})/\langle m_{max}\rangle$ is not very distinct. This is a manifestation of strong correlations in number of droplets of a given size. For example, if particles collide to form only two droplets of similar size, these two droplets may not collide for a very long time. Hence we observe large fluctuations even at $t = 2500$ s.~~

4 Fluctuations in conversion to rain drops and validity of the "constant SD" SDM

Fluctuations in time of conversion of cloud droplets to rain drops were studied using direct numerical simulations by ~~Onishi et al. (2015)~~Oni Following their notation, by $t_{10\%}$ we denote time after which 10% of mass of cloud droplets is turned into droplets with $r > 40$ µm. Droplets of this size should then quickly grow through coalescence. The time $t_{10\%}$ is used as a measure of efficiency of rain production. We will compare results of "the "one-to-one" " simulations with DNS and try to determine how many ~~super-droplets~~ SDs are needed in the ~~"constant SD"~~ "constant SD" simulations to accurately represent coalescence. The same initial droplet distribution and coalescence kernel is used as in Sec. 3.

In Fig. 4, values of mean ~~$\langle t_{10\%}\rangle$~~ $t_{10\%}$ for different initial number of droplets are plotted against the number of SDs. Results of both the ""one-to-one" " (rightmost points in each series) and ~~"constant SD"~~ the "constant SD" (rest of the points in the series) simulations are presented. ~~Horizontal line shows the result of solving~~ For comparison, $t_{10\%}$ obtained by solving the Smoluchowski equation using the ~~flux method from Bott (1998). The~~ "Bott algorithm is plotted (Bott, 1998). In the Bott algorithm, we used $\Delta t = 1$ s and mass bin spacing $m_{i+1} = 2^{1/10} m_i$. The same parameters were used in each simulation presented in this manuscript. Convergence tests were done for each case. The "one-to-one" " results converge with increasing system volume (i.e. increasing N_0) ~~. It is not clear if they would converge to Smoluchowski result, or to some higher value. The~~ to a value higher than the Smoluchowski result. The difference is probably caused by the numerical diffusion of the Bott algorithm. In the "constant SD" simulations, error caused by using SDs with $\xi > 1$ ~~(in "constant SD" simulations)~~ weakly

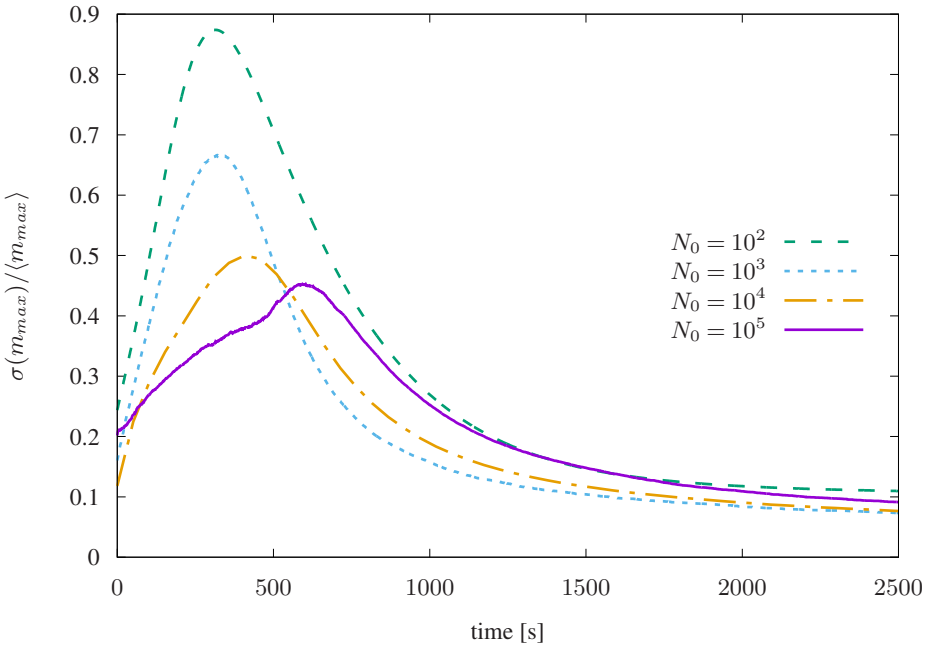


Figure 3. Relative standard deviation of mass of the largest droplet for different system sizes. Estimated from ensembles of $\Omega = 10^4$ "one-to-one" simulations for each value of N_0.

depends on the system size. Using 10^3 SDs gives $\langle t_{10\%}\rangle$ within 1% of the "~~"~~one-to-one~~"~~" value. Using 10^2 SDs causes about 10% ~~delay in rain formation. It is worth noting thaton modern computers, large eddy simulations (LES) with 10^2 SDs per cell are feasible, but those with $N_{SD} = 10^3$ would be very demanding~~difference. This shows that, in terms of computational cost, it is relatively cheap to obtain a good estimate of the average result of coalescence using the SDM. The SDM results are also compared with the results of DNS, in which air turbulence was not modelled, but hydrodynamic interactions between droplets were accounted for. We choose this kind of DNS, because it should be well described by the Hall kernel that is used in the SDM and in the Smoluchowski equation. It turns out that the Hall kernel gives too short autoconversion times. The same issue was observed by Onishi et al. (2015) (cf. Fig. 1(b) therein).

To analyze the amplification of fluctuations in the ~~"constant SD"~~ "constant SD" method, we plot the relative standard deviation of $t_{10\%}$ in Fig. 5. For reference, results ~~from DNS from Onishi et al. (2015)~~ of DNS from Onishi et al. (2015) are shown. Results from our "~~"~~one-to-one~~"~~" simulations are in good agreement with them. Small discrepancies are probably caused by the fact that ~~in Onishi et al. (2015) different coalescence kernels were used~~ the DNS included turbulence of various strength for different N_0. Results of "~~"~~the "one-to-one~~"~~" simulations were fitted with ~~a~~ the function $\alpha\sqrt{1/N_0}$ ~~with~~ , resulting in $\alpha = 6$. Figure 5 also presents fluctuations in ~~"constant SD"~~ the "constant SD" simulations for various N_{SD}. This type of simulations gives correct amplitude of fluctuations only for relatively low values of the ratio N_0/N_{SD}. For constant N_{SD}, as N_0 increases, the amplitude of fluctuations ~~correctly decreases~~ decreases correctly. Then, above some critical value of the N_0/N_{SD} ratio,

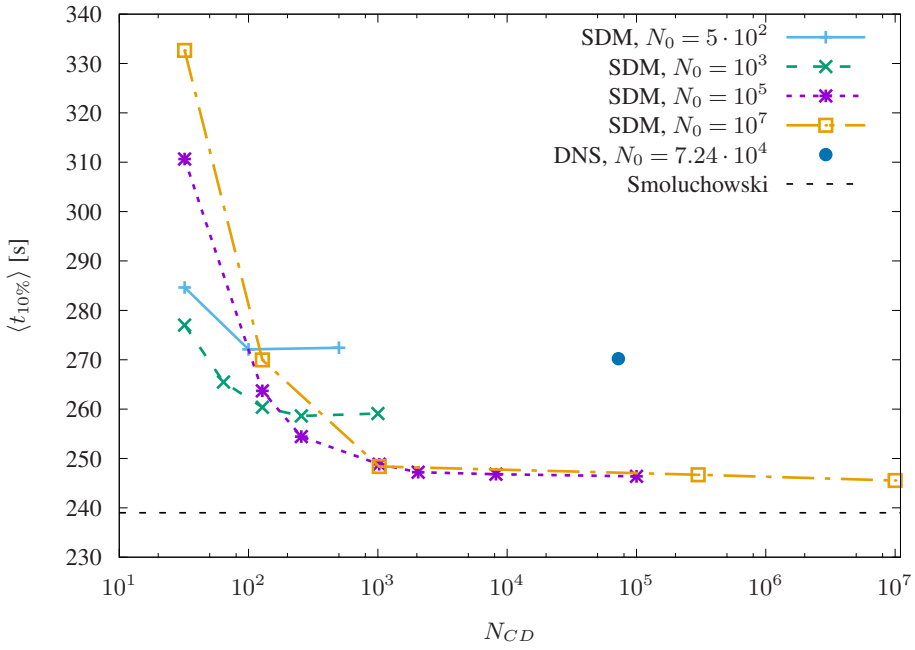


Figure 4. Mean $t_{10\%}$ for different system sizes and different numbers of computational droplets N_{CD}. In SDM simulations, $N_{CD} = N_{SD}$ and in DNS, $N_{CD} = N_0$. The single DNS result is taken from ~~ensemble~~ Onishi et al. (2015) (the NoT-HI case therein). Ensemble sizes are $\Omega \geq 10^3$ for SDM simulations and $\Omega = 10^2$ for DNS. ~~Errorbars~~ The 95% confidence intervals are smaller than plotted points. ~~Rightmost~~ The rightmost point ~~for~~ in each SDM series ~~is~~ comes from ~~"~~the "one-to-one~~"~~" simulations. Other points in SDM series are from the "constant SD" simulations with various values of N_{SD}. The horizontal line is ~~the~~ a value obtained ~~from~~ by numerically solving the Smoluchowski equation using the flux method from Bott (1998).

fluctuations stop to decrease and remain constant independent of the system size. This is a result of introducing unrealistic correlations between droplet sizes, a consequence of the low number of simulational particles (Bayewitz et al., 1974). We show the limiting, minimal value of relative standard deviation of $t_{10\%}$ in Fig. 6. It decreases as $\beta\sqrt{1/N_{SD}}$, with $\beta = 2$. By comparing it with $\alpha = 6$, we conclude that in order to obtain correct fluctuations in $t_{10\%}$, number of SDs has to be $N_{SD} \geq \frac{1}{9}N_0$. Using so many SDs is not feasible in ~~LES simulations~~Large Eddy Simulations (LES), but is possible in smaller scale simulations. Also, knowing α and β, we can estimate the magnitude of fluctuation amplification in the "constant SD" SDM.

5 Validity of the Smoluchowski equation

The Smoluchowski equation presents a mean-field description of the evolution of the size spectrum. It is exact only in the thermodynamic limit ($V \to \infty$). We will try to determine minimal system size for which the Smoluchowski equation can be used without introducing major errors. To do so, we analyze the evolution of θ, the ratio of rain water ($r \geq 40$ µm) content to

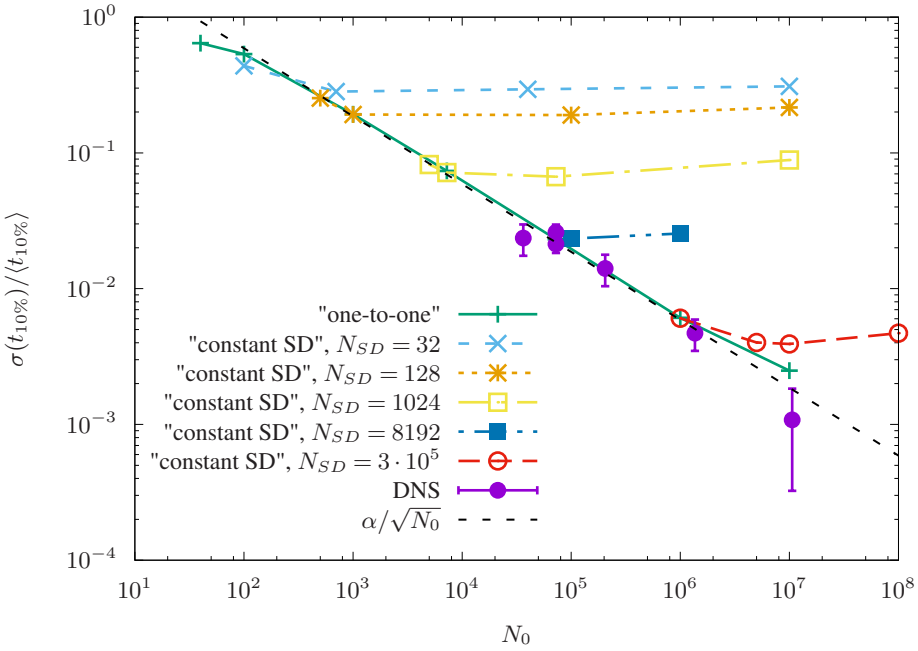


Figure 5. Relative standard deviation of $t_{10\%}$ against system size. ~~Our~~ SDM results are based on samples of ~~size of~~ at least 10^3 realizations. DNS results are taken from Onishi et al. (2015). Where not shown, errorbars are smaller than plotted points. ~~Curve fitting gave~~ The value $\alpha = 6$ was obtained by curve fitting to the "one-to-one" results.

the total water content. ~~Onishi et al. (2015)~~ Onishi et al. (2015) denote this value by τ. We do not adopt this notation to avoid confusion with the characteristic time.

We compare results of ~~"~~the "one-to-one~~" ~~~~"~~ simulations with solutions of the Smoluchowski equation ~~in~~ for two cases - with fast and ~~with~~ slow rain development. In both cases collision efficiencies for large droplets ~~were taken from Hall (1980)~~ are taken from Hall (1980) and for small droplets from ~~Davis (1972)~~Davis (1972). In the first case, we use the same initial distribution as in Secs. 3 and 4, which results in rapid rain development. As seen in Fig. 7, the Smoluchowski equation gives correct mean rain development rate for systems with $N_0 \geq 10^4$. ~~In smaller systems~~ The Smoluchowski curve is slightly shifted left, probably due to the numerical diffusion of the Bott algorithm, as discussed in Sec. 4. In systems smaller than $N_0 = 10^4$, rain develops slower than predicted by the Smoluchowski equation. Agreement of stochastic coalescence in large systems with the Smoluchowski equation for a similar initial distribution was shown in ~~See~~βelberg et al. (1996). ~~Onishi et al. (2015)~~ Seeβelberg et al. (1996). Onishi et al. (2015) present figures similar to Fig. 7, but obtained from DNS runs for $N_0 = 7.24 \cdot 10^4$ (~~Figs~~Fig. 1(~~a) and 1(~~b) therein). They also show good agreement between model results and the Smoluchowski equation, at least up to $t = 330$ s.

The second case is well below the size gap, i.e. the range of radii for which both collisional and condensational growths are slow. We use $\overline{r} = 9.3$ µm and $n_0 = 297$ cm^{-3} as in ~~Wang et al. (2006)~~Wang et al. (2006). In addition, we cut the distribution to 0 at $r = 20$ µm. This cutoff is used in the SDM modelling as well as when solving the Smoluchowski equation. That

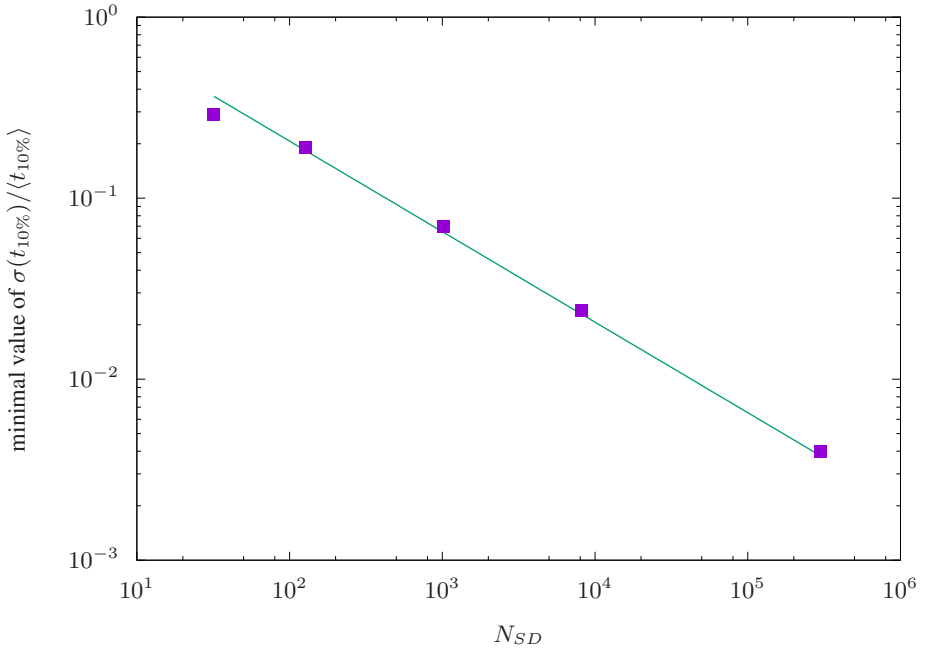


Figure 6. Minimal, limiting value of the relative standard deviation of $t_{10\%}$ for a given number of super-droplets (squares) in "constant SD" simulations. It is calculated as an average of the points to the right of the $\alpha/\sqrt{N_0}$ curve in Fig. 5. Line is the fitted function $\beta/\sqrt{N_{SD}}$ with $\beta = 2$.

way we get rid of the occasional very large ~~droplets~~ SDs present at $t = 0$ in some realizations ~~. Then~~of the SDM. For these initial conditions, rain development takes much longer and fluctuations can play a bigger role. Results are presented in Fig. 8. ~~Surprisingly, the "~~Again, we see convergence of the "one-to-one~~" results do not converge~~ " simulations to the Smoluchowski ~~results with increasing system size. We see convergence towards some other, higher value in~~ result, but in this case the system has to be larger ($N_0 \geq 10^7$) for the ~~time it takes for rain to form. Up to $\langle\theta\rangle = 0.2$ results for $N_0 = 10^5$ are the same as for $N_0 = 10^6$, but much different from the Smoluchowski results. The latter might be affected by numerical diffusion because the simulation runs for quite long before any rain is formed. For $\langle\theta\rangle \geq 0.2$, there are discrepancies between $N_0 = 10^5$ and $N_0 = 10^6$ results. The rate of growth of $\langle\theta\rangle$ decreases earlier with increasing system size~~Smoluchowski equation to be valid. The way how the "one-to-one" curves converge to the Smoluchowski curve is interesting. As in the first case, in smaller systems rain appears later than in larger systems. On the other hand, the rain formation rate (the slope of the curves in Fig. 8) in smaller systems starts to decrease at higher values of θ than in larger systems. In consequence, ~~after $t = 5000$ s we get lower rain fraction for $N_0 = 10^5$ than for $N_0 = 10^4$, and even lower for $N_0 = 10^6$.~~

~~As in Fig. 7, but for an initial distribution with $\bar{r} = 9.3$ µm, $n_0 = 297$ cm^{-3} and a cutoff at $r = 20$ µm.~~

~~The decrease of rain growth rate coincides with a decrease in~~ smaller systems can produce higher rain ratio than larger ones, although they started producing rain later (e.g. compare curves for $N_0 = 10^5$ and $N_0 = 10^7$ for $t > 4200$ s). The decrease in the

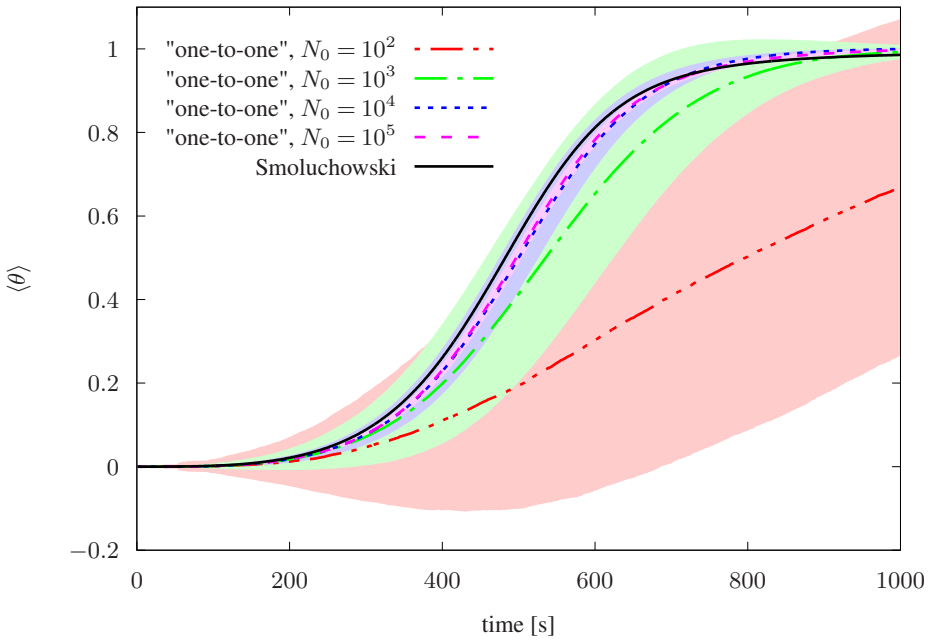


Figure 7. Rain content ratio θ for different system sizes averaged over ensembles of ~~$\Omega = 10^4$~~ $\Omega = 10^3$ simulations. Shaded regions show one standard deviation interval.

rain formation rate is associated with the decrease in the concentration of rain drops n_r, ~~as shown~~ plotted in Fig. 9. Number of rain drops decreases due to collisions between ~~droplets~~ drops from this category. A single ~~droplet that results out of~~ drop that is produced in such collision is less efficient at scavenging cloud droplets than the two pre-collision ~~droplets. As a~~ drops. In result, growth rate of θ decreases. Using large well-mixed volumes may introduce additional, unrealistic rain-rain collisions. Consider two droplets within a large cell that independently grow to the rain category. They have to be separated enough not to deplete liquid water from each other's surrounding as they grow. If we assume that the cell is well-mixed, they can immediately collide after becoming rain drops and generate an even larger drop. In reality, they could collide only after some time after becoming rain drops, because first they would need to overcome the initial separation. This means that using large well-mixed volumes, e.g. in the Smoluchowski equation, may result in underestimating the amount of rain produced.

In coalescence cells with $N_0 \leq 10^4$, we do not observe the decrease in the number of rain drops within 5000 s, probably because sizes of rain drops are similar. ~~For~~ In larger cells, more rain drops with a broader distribution are formed. In consequence, they collide more often, which decreases their number and the rate of collection of cloud droplets. It is likely that the ~~same effect is~~ additional rain-rain collisions in large volumes are responsible for the additional inflection point around $t = 500$ s in the plot of the relative standard deviation of the largest droplet mass for $N_0 = 10^5$ (cf. Fig. 3). ~~This~~ They could also lead to the deviation from the $\sim 1/\sqrt{N_0}$ scaling seen in Fig. 5. Fluctuations in cells with $N_0 = 10^7$ are greater than predicted using this scaling.

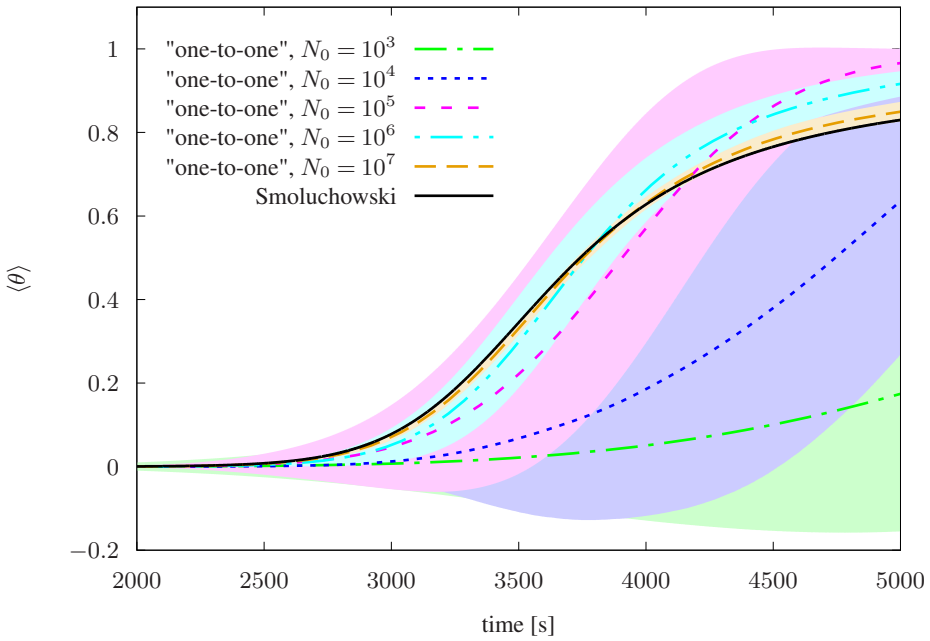


Figure 8. As in Fig. 7, but for an initial distribution with $\overline{r} = 9.3$ µm, $n_0 = 297$ cm^{-3} and a cutoff at $r = 20$ µm. The ensamble size is $\Omega = 10^8/N_0$.

~~Mean concentration of rain drops for the same initial distribution as in Fig. 8 from ensembles of at least $\Omega = 10^4$ simulations.~~

~~Judging from Fig. 9, we conclude that Smoluchowski equation consistently overestimates the number of rain drops during the initial phase, that is when $\langle n_r \rangle$ increases. Smoluchowski equation gives up to 50% higher values of $\langle n_r \rangle$.~~ We also observe that, although the amount of rain water depends strongly on ~~the~~ cell size, the number of rain drops does not. In larger cells rain drops acquire larger sizes through collisions with ~~rain drops~~cloud droplets, but rate of ~~production of~~ autoconversion of cloud droplets to rain drops is not affected ~~by using larger~~ much by cell size. ~~In box model simulations, the Smoluchowski equation produces too much rain if initial distribution is well below size gap and droplets slowly grow through coalescence. It is difficult to tell if using the Smoluchowski equation in cloud models overestimates the amount of rain. Possibly, condensational growth helps droplets cross the gap, leading to initial condition that is closer to the one in the first case ($\overline{r} = 15$ µm). For such initial condition, Smoluchowski equation gives correct results.~~

6 Lucky droplets

There is a well-established idea that some droplets undergo series of unlikely collisions and grow much faster than an average droplet (Telford, 1955; Scott, 1967; Marcus, 1968; Robertson, 1974; Mason, 2010). These few lucky droplets are

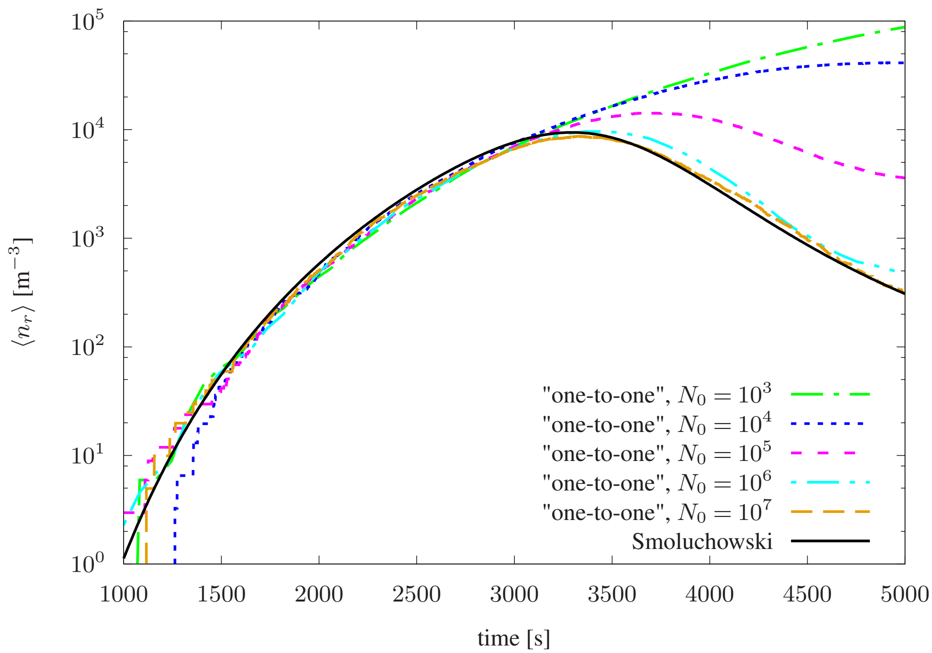


Figure 9. Mean concentration of rain drops from the same simulations as in Fig. 8.

Table 1. Average, standard deviation and sample size of time (in seconds) for the lucky realizations to produce single rain ~~droplet~~drop.

10^{-4}		$\gamma = 10^{-3}$			$\gamma = 10^{-2}$			$\gamma = 10^{-1}$			
~~$t_{40}^\gamma)$~~ $\sigma(t_{40})_\gamma$	$\gamma\Omega$	~~$\langle t_{40}^\gamma\rangle$~~ $\langle t_{40}\rangle_\gamma$	~~$\sigma(t_{40}^\gamma)$~~ $\sigma(t_{40})_\gamma$	$\gamma\Omega$	~~$\langle t_{40}^\gamma\rangle$~~ $\langle t_{40}\rangle_\gamma$	~~$\sigma(t_{40}^\gamma)$~~ $\sigma(t_{40})_\gamma$	$\gamma\Omega$	~~$\langle t_{40}^\gamma\rangle$~~ $\langle t_{40}\rangle_\gamma$	~~$\sigma(t_{40}^\gamma)$~~ $\sigma(t_{40})_\gamma$	$\gamma\Omega$	~~$\langle t_{40}^\gamma\rangle$~~ $\langle t_{40}\rangle_\gamma$
212	10	2930	356	10	4053	517	10^2	6365	1158	10^3	14777
120	10^2	1762	170	10^3	2400	267	10^4	3440	505	10^5	6500
173	3	1336	103	10	1717	176	10^2	2354	276	10^3	3912
33	2	1090	60	20	1334	85	200	1721	169	2000	2552
					1038	165	2	1301	176	20	1831

argued to be responsible for droplet spectra broadening and rain ~~formation~~ forming quicker than predicted by the Smoluchowski equation. Luck is supposed to be especially important during crossing of the size gap, when collisions happen rarely (Robertson, 1974; Kostinski and Shaw, 2005). A single droplet that would cross the size gap through lucky collisions could then initiate a cascade of collisions. ~~We use the same initial distribution as in the second case in Sec. 5. The mean radius is $\bar{r} = 9.3$ µm, well below the size gap.~~ Theoretical estimation of the ~~"luck factor"~~ “luck factor” was presented in ~~Kostinski and Shaw (2005)~~Kostinski and Shaw (2005). We use ~~"~~the “one-to-one~~"~~” simulations to test predictions from that paper.

We are interested in time t_{40} it takes for the largest droplet in the system to grow to $r = 40$ µm. We perform simulations for the same initial distribution as in the second case in Sec. 5. The mean radius is $\overline{r} = 9.3$ µm, well below the size gap. From an ensemble of Ω realizations, we select sub-ensembles of luckiest realizations, i.e. those with the smallest t_{40}. We consider sub-ensembles of size $\gamma\Omega$ with $log_{10}(\gamma) = -4, -3, -2, -1, 0$. In each sub-ensemble, we calculate ~~mean $\langle t_{40}^{\gamma} \rangle$ and standard deviation $\sigma(t_{40}^{\gamma})$~~the mean $\langle t_{40} \rangle_{\gamma}$ and the standard deviation $\sigma(t_{40})_{\gamma}$, where the subscript γ denotes the size of the sub-ensemble from which the statistic was calculated. The results for different cell sizes are shown in Tab. 1. There is a large variability in ~~$\langle t_{40}^{\gamma} \rangle$~~ $\langle t_{40} \rangle_{\gamma}$ with cell size. This is caused by the fact that t_{40} depends only on a single largest droplet. Larger cells contain more droplets, so probability of producing single large droplet increases with cell size. We notice that ~~$\langle t_{40}^{\gamma} \rangle$~~ $\langle t_{40} \rangle_{\gamma}$ is approximately the same along the diagonals of Tab. 1. For example, cell containing 10^6 droplets on average will produce first rain droplet in 30 minutes. If we divided it into 10 cells with 10^5 droplets each, the luckiest one would also produce a droplet in 30 minutes on average. This shows that using large coalescence cells does not affect formation of first rain drops. The differences discussed in previous Sections emerge later, when there are already some rain drops that can collide with each other. Moving to very small cells, we no longer observe same ~~$\langle t_{40}^{\gamma} \rangle$~~ $\langle t_{40} \rangle_{\gamma}$ along the diagonals. Ten cells with $N_0 = 10^2$ produce rain drops slower than a single cell with $N_0 = 10^3$. This is due to depletion of water droplets in small cells. The largest droplet a cell with $N_0 = 10^2$ can produce has $r \approx 43$ µm, close to the 40 µm rain threshold.

~~Kostinski and Shaw (2005)~~ Kostinski and Shaw (2005) estimate that the luckiest 10^{-3} fraction of droplets should cross the size gap around six times faster than average, while the luckiest 10^{-5} around nine times faster. We compare these values with our simulations for $N_0 = 10^3$. We choose this cell size, because it is the smallest one for which water depletion does not affect t_{40}. As far as t_{40} is concerned, larger cells behave exactly like an ensemble of cells of this size. We find ~~$\langle t_{40}^{10^{-3}} \rangle / \langle t_{40}^{1} \rangle \approx 3.7$ and $\langle t_{40}^{10^{-5}} \rangle / \langle t_{40}^{1} \rangle \approx 6$~~ $\langle t_{40} \rangle_{10^{-3}} / \langle t_{40} \rangle_{1} \approx 3.7$ and $\langle t_{40} \rangle_{10^{-5}} / \langle t_{40} \rangle_{1} \approx 6$. The value of ~~$\langle t_{40}^{10^{-5}} \rangle$~~ $\langle t_{40} \rangle_{10^{-5}}$ was estimated at 1090 s based on values along the diagonal for larger γ and larger N_0. These ratios are lower than given in ~~Kostinski and Shaw (2005)~~Kostinski and Shaw (2005), showing that their theoretical analysis ~~overestimates the "luckiness"~~ slightly overestimates the "luckiness" in droplet growth. ~~Nevertheless, we agree with their conclusion that fluctuations play an important role in rain formation. Thanks to lucky collisions in some realizations (or, alternatively, in some parts of the cloud), mean concentration of rain drops after 30 minutes is about 200 m^{-3}. On the other hand, using the Smoluchowski equation leads to higher rain drop concentration than can be produced by lucky collisions (cf. Fig. 9). Significant role of fluctuations can also be seen in Fig. 8. Relative standard deviation of θ is high in small cells ($N_0 \leq 10^4$). This implies that small parts of the cloud could produce significant amount of rain much faster than average.~~

7 ~~Coalescence~~ Size of a well-mixed coalescence cell~~size~~

In the previous ~~sections~~ Sections we have seen that size of the coalescence cell has a profound impact on the ~~system evolution. Possibly, many of these differences would disappear once turbulent droplet motion and sedimentation are modeled~~evolution of the system. In this Section we ~~try to determine~~ estimate the size of a cell that ~~could be used in such modeling~~can be assumed to be well-mixed. All methods in which ~~each droplet within a cell can collide with any other droplet within the same cell~~

probability of collision of droplets depends only on the instantaneous state of the cell and not on its history rely on the assumption that the cell is well-mixed. This includes the master equation, SSA, SDM as well as the Smoluchowski equation. The assumption that a cell is well-mixed is valid if $\tau_{mix} \ll \tau_{coal}$, where τ_{coal} and τ_{mix} are the characteristic times for coalescence and cell homogenization, respectively (Lehmann et al., 2009; Gillespie et al., 2014). By well-mixed we mean that droplets should be distributed homogeneously within the cell before every collision. Droplet coalescence generates inhomogeneities, i.e. correlations between droplet positions and sizes. ~~Consider two droplets growing independently within a cell. After gaining large sizes, they collide and generate even larger droplet. In reality they could not both obtain large sizes before the collision, because they would deplete liquid water from each other's surrounding.~~

Rigorously, characteristic time for coalescence is the mean time between coalescence events, as in diffusion-limited chemical systems (Gillespie et al., 2014). To estimate its magnitude, consider a single large collector droplet falling through a field of smaller droplets. Using geometric coalescence kernel with efficiency E, the mean time between collisions is $\tau_{coal} = (E\pi(r_l + r_s)^2 v_r n_s)^{-1}$, where r_l and r_s are radii of large and small droplets, v_r is the relative velocity and n_s is the concentration of small droplets. For $r_l = 100$ µm, $r_s = 10$ µm, $v_r = 70$ cm/s, $E = 1$ and $n_s = 100$ cm^{-3} we get $\tau_{coal} \approx 0.4$ s.

Droplets in the cell can be mixed through turbulence. Turbulence acts similarly to diffusion and its characteristic time for mixing is $\tau^t_{mix} = (V^{(2/3)}/\varepsilon)^{(1/3)}$, where V is cell volume and ε is turbulent energy dissipation rate (Lehmann et al., 2009). Turbulent energy dissipation rate in clouds is in the range from 10 cm^2/s^3 for stratocumulus clouds to 10^3 cm^2/s^3 for cumulonimbus clouds (Malinowski et al., 2013; Grabowski and Wang, 2013). Let us assume that $\tau^t_{mix} \ll \tau_{coal}$ is satisfied if $\tau^t_{mix} = 0.1\tau_{coal}$. Even in the most turbulent clouds, this means that the coalescence cell has to be very small $V \approx 1.5{\cdot}10^{-2}$cm^3. On average, this volume would contain around one droplet, depending on concentration of droplets. ~~The Smoluchowski equation cannot be used for such small populations. Using super-droplet microphysics~~ For such small coalescence volumes, the Smoluchowski is not valid and SDM would be very cumbersome, because extremely short time steps would be required~~to decouple motion from collisions~~. To use larger cells, we need to choose some less strict value of characteristic time of coalescence. ~~For instance, Shima et al. (2009) assume $\tau_{coal} = 100$ s without much explanation.~~ Some larger cell size, that would be approximately well-mixed, could be found phenomenologically through ~~exact~~ fine-grid simulations including droplet motion. One example of such reference simulations are DNS runs from ~~Onishi et al. (2015)~~ Onishi et al. (2015) discussed in Sec. 5. They prove that in the case with $\overline{r} = 15$ µm, the Smoluchowski equation gives correct results. This suggests that cells with $N_0 \geq 10^4$ can be used in this case.

Another process that can mix droplets is sedimentation. It is difficult to assess its timescale, because it strongly depends on droplet sizes. Droplets of similar sizes are not mixed by sedimentation, but it is efficient at mixing rain drops with cloud droplets. We can expect that it would prevent depletion of cloud droplets in the surrounding of a rain droplet that was observed for smallest cells in Secs. 3 and 6. Sedimentation acts only in one direction, so it could only allow us to use cells larger only in the vertical direction.

8 Conclusions

The super-droplet method can exactly represent stochastic coalescence in a well-mixed volume. It was compared with the master equation approach (see Sec. 3) and with direct numerical simulations (see Sec. 4). Precision of the SDM is controlled by the number of super-droplets used. Fluctuations in the autoconversion time are represented well if $N_{SD} \geq N_0/9$. Using smaller N_{SD} increases standard deviation of autoconversion time by a factor $\frac{1}{3}\sqrt{N_0/N_{SD}}$ (cf. Sec. 4). It is computationally less expensive to correctly reproduce mean autoconversion time. Using $N_{SD} = 10^3$ gives mean results within a 1% margin, while using $N_{SD} = 10^2$ - within 10%.

The SDM was used to study stochastic coalescence for two initial droplet size distributions - with small ($\overline{r} = 9.3$ µm) and with large ($\overline{r} = 15$ µm) droplets. They result in slow and fast rain formation, respectively. Dependence of the system behavior on ~~coalescence cell size~~ size of the well-mixed coalescence cell was observed, especially in the small droplets case. Cell size not only affects fluctuations in the observables, but also their expected ~~value~~values. If the coalescence cell is small~~($N_0 < 10^3$)~~, sizes of droplets are strongly correlated and depletion of cloud water plays an important role. In ~~reality~~real clouds, these two effects are probably not manifested, because collector drop sedimentation acts against them. ~~If the coalescence cell is relatively large ($N_0 > 10^4$)~~In relatively large cells, rain drops ~~that in reality would form far from each other and would need time to get close, can collide immediately. This is because the coalescence cell is assumed to be well-mixed, which is usually not true. We estimate a~~ collide with each other more often than in small cells. This leads to a reduction in the rate of conversion of cloud water to rain water, because scavenging of cloud droplets becomes less efficient. In consequence, highest rain content is produced in cells of intermediate sizes. Possibly, these additional rain-rain collisions can be justified by turbulent droplet motion and sedimentation, but they also might be an artefact caused by using an unrealisticaly large well-mixed ~~(with respect to coalescence) volume in the most turbulent clouds to be only $1.5 \cdot 10^{-2}$ cm^3.~~

~~Unrealistic collisions between rain drops, caused by the assumption that coalescence cell is~~ volume. Fine-grid computer modeling with explicit droplet motion could be used to resolve this issue. If the additional collisions were found to be unrealistic, it would mean that cloud models that use large well-mixed ~~,~~cells, e.g. by using the Smoluchowski equaton, produce too little rain.

The additional rain-rain collisions do not affect results if droplets are initially large. Then, collisions of cloud and rain drops and between cloud droplets are frequent, so ~~relatively rare~~ increase in the rate of collisions between rain drops ~~are~~ is not important. The mean behavior of the system converges to the Smoluchowski equation results with increasing system size. Good agreement with it is found for systems with $N_0 \geq 10^4$. The picture is different if droplets are initially small. Conversion of cloud droplets into rain drops is slow, so the decrease in rain drop concentration due to ~~these unrealistic~~ the additional collisions is relatively more important. ~~Coalescence of rain drops decreases the rate of collection of cloud droplets, because a single larger drop has smaller collisional cross section than two smaller drops with the same total volume. In consequence, mean behavior of the system no longer converges with increasing cell size. Up to $\langle\theta\rangle = 0.2$, results for $N_0 = 10^5$ are the same as for $N_0 = 10^6$. Then, rate of rain growth decreases sooner in the larger coalescence cell.~~ The Smoluchowski equation is found to be valid for $N_0 \geq 10^7$ for the slow-coalescence case. One could expect that condensational growth leads to initial conditions

with high radii of droplets, for which the additional collisions are not important. Li et al. (2017) have shown that condensation can regulate differences between Eulerian and Lagrangian coalescence schemes. Discrepancies between these schemes that they observed in simulations with condensation and coalescence were smaller than in pure coalescence simulations.

Another aspect of the slow-coalescence scenario is that in it, some lucky droplets can grow much faster than average droplets. We found that a single luckiest droplet out of a thousand grows 3.7 times faster than average and the luckiest out of a hundred thousand - 6 times faster. These values are ~~smaller than predicted by Kostinski and Shaw (2005), but large enough to be important for quick formation of rain~~slightly smaller than the estimation of Kostinski and Shaw (2005).

~~The size of~~ We estimate a well-mixed ~~volume, i.e. a volume within which droplets are randomly rearranged through turbulence between coalescence events,~~ (with respect to coalescence) volume in the most turbulent clouds to be only $1.5 \cdot 10^{-2}$ cm^3. It is of the order of the volume occupied by a single droplet. Larger cells can be assumed to be only approximately well-mixed. For example, in the fast-coalescence case, DNS modeling gives the same results as the Smoluchowski equation (Onishi et al., 2015). Box model simulations using well-mixed volume with ~~$N_0 = 10^5$~~ $N_0 = 10^4$ droplets also gives the same results. Therefore it can be assumed that such volume is approximately well-mixed in the case of fast coalescence. ~~On the other hand, in~~ In the slow-coalescence case, ~~box model simulations do no converge to~~ the ~~Smoluchowski result. This implies that models that use~~ well-mixed volume needs to be larger than in the fast-coalescence case for the Smoluchowski equation ~~might produce rain too soon. The real behavior of the system could be determined through DNS modeling or SDM simulations with droplet motion.~~

to be valid. Size of an approximately well-mixed cell for this case can be determined using DNS with initially small droplets. Cells used in LES are typically ten orders of magnitude larger than a well-mixed volume. They do not necessarily have to be well-mixed. It is sufficient if they are homogeneous, i.e. they are an ensemble of identical, approximately well-mixed sub-cells. Some statistical moments for such ensembles were presented in this work. In general, it is not clear what could be the size of these sub-cells and if the Smoluchowski equation is valid for them. ~~We have shown that for initially small droplets, the Smoluchowski equation gives wrong results, but is correct for initially large droplets. One could hope that condensational growth leads to initial conditions close to the ones for which the Smoluchowski equation is valid, but justifying it would require further research.~~

9 Code availability

Simulation code is available at https://github.com/pdziekan/coal_fluctu. The libcloudph++ library is available at https://github.com/igfuw/lib

Acknowledgements. This study was financed from Poland's National Science Center ~~"~~"POLONEZ 1~~"~~" grant 2015/19/P/ST10/02596 (this project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 665778) and Poland's National Science Center ~~"~~"HARMONIA 3~~"~~" grant 2012/06/M/ST10/00434. Numerical simulations were carried out at the Cyfronet AGH computer center, accessed through the PLGrid ~~project~~portal. We are grateful to Wojciech W. Grabowski for fruitful discussions.

References

Alfonso, L. (2015). An algorithm for the numerical solution of the multivariate master equation for stochastic coalescence. *Atmospheric Chemistry and Physics*, 15(21):12315–12326.

Alfonso, L. and Raga, G. B. (2016). The impact of fluctuations and correlations in droplet growth by collision-coalescence revisited. part i: Numerical calculation of post-gel droplet size distribution. *Atmospheric Chemistry and Physics Discussions*, 2016:1–33.

Alfonso, L. and Raga, G. B. (2017). The impact of fluctuations and correlations in droplet growth by collision–coalescence revisited – part 1: Numerical calculation of post-gel droplet size distribution. *Atmospheric Chemistry and Physics*, 17(11):6895–6905.

Andrejczuk, M., Reisner, J., Henson, B., Dubey, M., and Jeffery, C. (2008). The potential impacts of pollution on a nondrizzling stratus deck: Does aerosol number matter more than type? *Journal of Geophysical Research: Atmospheres*, 113(D19).

Arabas, S., Jaruga, A., Pawlowska, H., and Grabowski, W. (2015). libcloudph++ 1.0: a single-moment bulk, double-moment bulk, and particle-based warm-rain microphysics library in c++. *Geoscientific Model Development*, 8(6):1677–1707.

Bayewitz, M. H., Yerushalmi, J., Katz, S., and Shinnar, R. (1974). The extent of correlations in a stochastic coalescence process. *Journal of the Atmospheric Sciences*, 31(6):1604–1614.

Beard, K. (1976). Terminal velocity and shape of cloud and precipitation drops aloft. *Journal of the Atmospheric Sciences*, 33(5):851–864.

Bird, G. (1994). *Molecular gas dynamics and the direct simulation monte carlo of gas flows*. Clarendon, Oxford.

Bott, A. (1998). A flux method for the numerical solution of the stochastic collection equation. *Journal of the atmospheric sciences*, 55(13):2284–2293.

Davis, M. H. (1972). Collisions of small cloud droplets: Gas kinetic effects. *Journal of the Atmospheric Sciences*, 29(5):911–915.

Gillespie, D. T. (1972). The stochastic coalescence model for cloud droplet growth. *Journal of the Atmospheric Sciences*, 29(8):1496–1510.

Gillespie, D. T. (1975). An exact method for numerically simulating the stochastic coalescence process in a cloud. *Journal of the Atmospheric Sciences*, 32(10):1977–1989.

Gillespie, D. T., Petzold, L. R., and Seitaridou, E. (2014). Validity conditions for stochastic chemical kinetics in diffusion-limited systems. *The Journal of chemical physics*, 140(5):02B604_1.

Grabowski, W. W. and Wang, L.-P. (2013). Growth of cloud droplets in a turbulent environment. *Annual Review of Fluid Mechanics*, 45:293–324.

Hall, W. D. (1980). A detailed microphysical model within a two-dimensional dynamic framework: Model description and preliminary results. *Journal of the Atmospheric Sciences*, 37(11):2486–2507.

Kostinski, A. B. and Shaw, R. A. (2005). Fluctuations and luck in droplet growth by coalescence. *Bulletin of the American Meteorological Society*, 86(2):235–244.

Lehmann, K., Siebert, H., and Shaw, R. A. (2009). Homogeneous and inhomogeneous mixing in cumulus clouds: Dependence on local turbulence structure. *Journal of the Atmospheric Sciences*, 66(12):3641–3659.

Leyvraz, F. (2003). Scaling theory and exactly solved models in the kinetics of irreversible aggregation. *Physics Reports*, 383(2):95–212.

Li, X.-Y., Brandenburg, A., Haugen, N. E. L., and Svensson, G. (2017). Eulerian and lagrangian approaches to multidimensional condensation and collection. *Journal of Advances in Modeling Earth Systems*, pages n/a–n/a.

Malinowski, S., Gerber, H., Plante, J.-L., Kopec, M., Kumala, W., Nurowska, K., Chuang, P., Khelif, D., Haman, K., et al. (2013). Physics of stratocumulus top (post): turbulent mixing across capping inversion. *Atmospheric Chemistry and Physics*, 13(24):12171–12186.

Marcus, A. H. (1968). Stochastic coalescence. *Technometrics*, 10(1):133–143.

Mason, B. J. (2010). *Physics of clouds*. Clarendon Press.

Onishi, R., Matsuda, K., and Takahashi, K. (2015). Lagrangian tracking simulation of droplet growth in turbulence–turbulence enhancement of autoconversion rate. *Journal of the Atmospheric Sciences*, 72(7):2591–2607.

Riechelmann, T., Noh, Y., and Raasch, S. (2012). A new method for large-eddy simulations of clouds with lagrangian droplets including the effects of turbulent collision. *New Journal of Physics*, 14(6):065008.

Robertson, D. (1974). Monte carlo simulations of drop growth by accretion. *Journal of the Atmospheric Sciences*, 31(5):1344–1350.

Scott, W. T. (1967). Poisson statistics in distributions of coalescing droplets. *Journal of the Atmospheric Sciences*, 24(2):221–225.

Seeβelberg, M., Trautmann, T., and Thorn, M. (1996). Stochastic simulations as a benchmark for mathematical methods solving the coalescence equation. *Atmospheric Research*, 40(1):33–48.

Shima, S.-i., Kusano, K., Kawano, A., Sugiyama, T., and Kawahara, S. (2009). The super-droplet method for the numerical simulation of clouds and precipitation: A particle-based and probabilistic microphysics model coupled with a non-hydrostatic model. *Quarterly Journal of the Royal Meteorological Society*, 135(642):1307–1320.

Smoluchowski, M. V. (1916). Drei Vortrage uber Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen. *Zeitschrift fur Physik*, 17:557–585.

Tanaka, H. and Nakazawa, K. (1993). Stochastic coagulation equation and validity of the statistical coagulation equation. *Journal of geomagnetism and geoelectricity*, 45(5):361–381.

Telford, J. (1955). A new aspect of coalescence theory. *Journal of Meteorology*, 12(5):436–444.

Unterstrasser, S., Hoffmann, F., and Lerch, M. (2016). Collection/aggregation algorithms in lagrangian cloud microphysical models: Rigorous evaluation in box model simulations. *Geoscientific Model Development Discussions*, 2016:1–49.

Wang, L.-P., Xue, Y., Ayala, O., and Grabowski, W. W. (2006). Effects of stochastic coalescence and air turbulence on the size distribution of cloud droplets. *Atmospheric Research*, 82(1):416–432.