

Response to the reviews

September 20, 2017

We would like to express our gratitude to the reviewers for the time they invested in improving the paper. We have made changes to the manuscript in order to make the presentation more clear. Also, following the editor's advice, we have consulted a researcher from outside of our field on readability of the paper.

Answer to the Anonymous Referee #2.

I appreciate very much the improvements made by the authors. The scientific content of the paper is novel and original. Nevertheless I believe that the presentation should be still improved in order to help future readers and to make sure it draws the attention it deserves. I still have the feeling that your presentation should be more explicit and that you should guide the readers better through the paper.

Main points

The newly added part towards the end of section 1 is a first step, but each section should start with a motivation and explain the goals behind the next tests. This is already nicely done in Section 4, 5, 6 and 7. On some occasions it may suffice to simply change the order of the presented material.

Sections 2 and 3 now start with short introductions that explain their goals.

Some examples: 1. Whereas the intention of showing the size distribution in Fig.1 is clear to everyone and needs no further explanation, most readers probably do not know beforehand why you analyse of $\sigma(m_{\max})/E(m_{\max})$ and that this quantity is related to the gelation time. Your sentence on page 5, line 30 is essential in motivating what you do and hence should appear earlier.

The sentence now appears before we discuss Fig.2.

2. You may still expand the description of the algorithm to make your paper self-contained such that readers are not forced to read Shima. Your whole paper is based on this algorithm, so I think it is worth investing a few more lines. It probably suffices to present a condensed version of their Eqs. 12-19, better explain multiple collections, and stress the point that only integer multiplicities are allowed. It may also help to highlight that the constant N_{SD} simulation with $N_{SD}=N_0$ is different to a one-to-one simulation with N_0 . Still I think that some formulations are not explicit enough. See e.g. my comment from the last review you should write that one SD is created... Contrary to your response, I do not think you already say it. You just say that the radius is randomly selected. Better write that one SIP per bin is created....

We agree that it will be easier for the readers if they do not have to go through all of the Shima's paper before reading our work. Section 2 has been considerably expanded following the suggestions. We believe that it now includes all necessary information on how coalescence is done in the super-droplet method.

3. The following comment nicely demonstrates your implicit style of writing: The number of used realizations must be mentioned in the text, not only in the figure caption. In section 3, even the fact you analyse a certain number of realizations is not really mentioned directly. It is only implicitly clear by saying average or because you analyse sigmas.

It is true that it may have been stated more clearly. We have made an effort to give all necessary information more explicitly throughout the paper.

4. Could you expand the description of Table 1 in the text? I do not understand the meaning of the third column. Hence, I am not able to reconstruct all parameters of the individual cases. In particular I'm confused about values like 2, 3, 200 etc.

We have expanded the caption of Table 1.

Discussion points

I was wrong in stating that all SD methods solve the KCE (see points 4 and 5 of original review) and you convinced me that the probabilistic nature of the all-or-nothing approach has the same source of variability as the master equation. However, this is not true for the

Riechelmann and Andrejczuk algorithms. They are not probabilistic (no Monte-Carlo approach is used, instead they solve the average KCE) and behave differently in the limiting case (explanation follows in the next paragraph). So I would put it the following way: Some SD methods are based on KCE, some are based on the master equation. Hence, to avoid confusion, your statements throughout the manuscript should be reformulated, as they do not hold for SD methods in general.

We agree that some SD methods are based on the KCE. It was not our intention to give any conclusions on SD methods in general, only on the method of Shima. The misconception might be caused by the fact that there is no standard nomenclature in use. To clarify the text, now in the introduction we explicitly say that by the SD method we mean the method of Shima and that our conclusions are valid only for this method. In addition, we added a reference to the Shima method in the abstract.

In your and Shimas application of the all or nothing algorithm the multiplicities are integer values. Due to the design of the algorithm multiplicities remain integer, if integer values are used at initialisation. This is different in the Riechelmann and Andrejczuk algorithms, they produce real numbers, even for integer initialisations. So the limiting case of a one-to-one simulation does not reduce to the master equation. Hence, my impression is that the various SD methods are not equivalent in the limiting case. Note that the all-or-nothing algorithm can also be applied with real numbers (see Unterstrasser et al, 2017).

The finding in Sec 4 $N_{SD} > 1/9 N_0$ has strong implications on the feasibility of LES. May it be possible that with a full sampling of the SD pairs the constraint on N_{SD} is less strict? You may add at least a full sampling line for $N_{SD}=32$ in Fig.5 to get a rough tendency. I acknowledge the tests you show in Figs.1&2, but those may not suffice to prove the equivalence between the full and linear sampling for all applications.

It is true that LES with $N_{SD} > 1/9 N_0$ is not feasible. Using less SDs results in high fluctuations, but it does not mean that such simulations are useless. We have done LES of a stratocumulus with $N_{SD}=100$ and the resulting time series and averaged profiles are in qualitative agreement with results of LES with the KCE.

As suggested by the Reviewer, we have added a line for $N_{SD}=32$ with full sampling to Fig. 5. It is in good agreement with the line for $N_{SD}=32$ with linear sampling, so we conclude that the linear sampling does not affect the constraint on the number of SDs.

Minor Points:

You define SSA as the algorithm by Gillespie, but later on SSA refers to the algorithm by Alfonso, doesn't it?

SSA is the algorithm by Gillespie and it was used by Alfonso in the paper in which he introduces his own method. We now say it more clearly in Section 3.

p.2., l.24: You miss to cite the Lagrangian cloud model by Slch & Krcher,2010. By the way it uses also the all-or-nothing approach for particle collisions.

We added the citation.

p.3, l.20: I would say that only the opposite direction is true, i.e. all droplets in a SD are identical, but not all identical droplets are necessarily represented by one SD. You can well have two or ten SDs that represent all those droplets with similar properties.

We agree. The sentence now says that a SD represents *many*, not *all*, droplets with similar properties.

p.4, l.24: you may add which can only happen if two SDs with identical eta collide.

Done.

p.6, l.7: you probably refer to the exponential distribution used in Section 3. Check also for other occurrences.

Yes, we refer to the distribution which is used in Section 3 to initialize droplet sizes. We think that it is clearly stated. We write that the *initial* droplet distribution is the same to avoid any misunderstanding that the droplet distribution after the simulation is the same as in Section 3.

To make the connection between Fig. 5 and 6 clearer, you may use the same colours for the squares

We did so.

in Fig 6 as in Fig.5 p.8, l.1: simulational = computational?

Changed.

p.9, l.6: How can it be that DNS results agree well with KCE results, even though in Fig. 4 you argue that the autoconversion time of KCE (and SDM) is too short?

That is a good point. The KCE results agree with DNS results only if the Long kernel is used in the KCE, what is now stated in Section 5. If the Hall kernel is used, the autoconversion time of the KCE is too short. We conclude that it is not the KCE approach that is invalid, but the kernel used.

p.10, l.13: I thought the conclusion would be that rain production is overestimated. What do you mean with amount of rain, mass or number of rain drops? Maybe it is the case, that the mass is overestimated and the number is underestimated?

We meant that the mass of rain is underestimated. Now we say explicitly that both the number of rain drops and the mass of rain are underestimated.

p.12, l.3: Do you want to say ..it takes until the first droplet grows to $r=40\mu\text{m}$? Your formulation could cause confusion, as it changes over time, which droplet is the largest.

We have changed that following the suggestion.

p.13,l14 and l15: 1.) the ratios must be flipped to get numbers >1 .

The error has been corrected.

2.) Did Kostinski & Shaw results depend on system size? Your results do: For $N_0=10^4$ and 10^5 , the ratio goes down from 3.7 to 2.9 ($=3912/1336$) and 2.3 ($=2552/1090$). Whats the interpretation of this?

That is a very good point and led us to change our conclusions - now we conclude that our results are in agreement with the Kostinski & Shaw theory. By calculating the ratios of t_{40} we want to find out how much faster the luckiest droplet can grow compared to an average droplet. From the fact that values of t_{40} are similar along the diagonals of Table 1 we concluded that, when looking for the time for the first droplet to grow to $r=40$ microns, large cells behave like ensembles of cells with $N_0=10^3$. We used to think that smaller cells distort the result, so we were calculating the "luck factor", i.e. the ratio of lucky t_{40} to average t_{40} , for cells with $N_0=10^3$. The values of "luck factor" calculated using results for larger cells are smaller, because each larger cell itself is an ensemble of smaller cells, so it is an ensemble of droplets that can independently grow to $r=40$ microns.

The text and our reasoning has been changed in the following way. To better show that larger cells behave like an ensemble of smaller cells, we added a plot of $\langle t_{40} \rangle$ vs N_0 / γ (Fig. 10). Based on this plot we conclude that even smaller cells, with $N_0=10^2$, do not distort the results. Such cells can produce only a single drop with $r=40$ microns, so calculating the “luck factor” using these cells does in fact mean comparing single droplets and not ensembles of droplets. We find that the “luck factor” calculated from results for $N_0=10^2$ is in agreement with the Kostinski & Shaw theory, so we conclude that their theory is correct.

Typos

p.5, l.15: of discrepancies

**p.6, l.12: each simulation -> any Bott simulation
ensemble, not ensamble**

The reference list contains several small errors. I guess this is mostly due to the fact that in your bib file the paper titles are not embraced by title. Then all words appear in lower case, see Alfonso & Raga, Li, Malinowski, ..

Unterstrasser should be cited with the GMD, not the GMDD article.

Typos have been 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 of the collisional growth of cloud droplets is studied ~~in a box model~~ using the super-droplet method (SDM) ~~of Shima et al. (2009). Statistics are calculated from ensembles of simulations of collision-coalescence in a single well-mixed cell.~~ The SDM is compared with direct numerical simulations and the master equation. It is argued that SDM simulations in which one ~~super-droplet~~ computational droplet represents one real droplet are at the same level of precision as the master equation. Such simulations are used to study fluctuations in the autoconversion time, the sol-gel transition and the growth rate of lucky droplets, 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, collisions between rain drops are more frequent and ~~can also~~ this also can 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~~ The 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 multiplicities 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 ~~$V_{mix} = 1.5 \cdot 10^{-2}$~~ $V_{mix} = 1.5 \cdot 10^{-2}$ cm³. The Smoluchowski equation is not valid in such small volumes. It is argued that larger volumes can be considered approximately well-mixed, but such approximation needs to be supported by a comparison with fine-grid simulations 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 between numbers~~ 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.

5 ~~The-~~

~~Moreover, although the~~ Smoluchowski equation can be written for the discrete number of droplets of given size, ~~but often droplet concentration is used instead~~ it is more often used for droplet concentrations. This adds an additional assumption that the coalescence volume is large, somewhat in agreement with ~~neglecting fluctuations~~ neglecting fluctuations in number of collisions and correlations in droplet numbers (Gillespie, 1972).

10 A number of methods alternative to the Smoluchowski equation exist. They are capable of ~~solving~~ addressing 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~~ DNS, trajectories of droplets are ~~modeled~~ simulated 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
15 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 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
20 equation was analytically solved only for 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), but cannot be solved for any realistic coalescence kernel. Solving the master equation numerically is extremely difficult due to a huge phase space to be considered. Recently, 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, 2017). Alternatively, the stochastic
25 simulation algorithm (SSA) (Gillespie, 1975; Seesselberg et al., 1996) can be used to model single trajectory obeying the master equation, but obtaining large enough statistics would require very long computations.

Several Lagrangian methods have been developed to model cloud microphysics (~~Andrejczuk et al., 2008; Shima et al., 2009; Riechelmann et al., 2010~~). 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. We will refer to these computational particles as super-droplets (SDs). The
30 words “droplets” and “drops” are reserved for real hydrometeors. 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 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. Throughout the paper, by the name “super-droplet method” we refer to this specific algorithm and any conclusions
35 regarding SDM are valid only for the Shima SDM. The Shima et al. (2009) algorithm is not based on the Smoluchowski equa-

tion, 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. 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-cell 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 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-a cell has to be for the Smoluchowski equation to correctly represent the rate of rain formation. Throughout the paper we observe that evolution of the system-droplet size spectrum strongly depends on its size-size of the coalescence cell. The size of a well-mixed air parcel-volume is estimated in Sec. 7 and some implications for cloud simulations are discussed in Sec. 8.

2 The super-droplet method

In this Section we present how collision-coalescence is handled in the super-droplet method. Further information about SDM can be found in Shima et al. (2009). 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 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 the 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 and K is the coalescence kernel and V is the volume of the box. 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-many 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 within a single cell, 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. Only integer multiplicities are allowed. In the algorithm of Shima et al. (2009), two simplifications are made that may affect the amplitude of fluctuations in the number of collisions. The first simplification is that SDs collide in an “all-or-nothing” manner. If a collision happens, each real droplet represented by the SD with lower multiplicity collides with a single droplet represented-represented by the SD with higher multiplicity. If the i -th and j -th SDs

collide, their parameters are updated to:

$$\begin{aligned}\xi_j' &= \xi_j, & \xi_i' &= \xi_i - \xi_j \\ r_j' &= (r_i^3 + r_j^3)^{1/3}, & r_i' &= r_i,\end{aligned}\tag{1}$$

where primes denote post-collisional values and we assume $\xi_j \leq \xi_i$. Intuitively, one would expect ~~that this should be~~ the “all-or-nothing” procedure to lead to larger fluctuations than in ~~the a~~ real system, because the number of collision trials is artificially reduced.

- 5 The second simplification, that we will refer to as “linear sampling”, is that instead of considering all $N_{SD}(N_{SD}-1)/2$ $N(N-1)/2$ collision pairs, only $\lfloor N_{SD}/2 \rfloor \lfloor N/2 \rfloor$ non-overlapping pairs are randomly selected. $N_{SD}-N$ is the number of SDs in the coalescence volume cell 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. Probability of coalescence of two SDs i and j that belong to the same collision pair is $P_{SD}(r_i, r_j, \xi_i, \xi_j) = \max(\xi_i, \xi_j) P(r_i, r_j) (N_{SD}(N_{SD}-1)/2) / \lfloor N_{SD}/2 \rfloor$ $(\xi_i, \xi_j) P(r_i, r_j) (N(N-1)/2) / \lfloor N/2 \rfloor$ (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 perform two types of simulations. In the “one-to-one” simulations, all SDs have multiplicity $\xi = 1$. That way the “all-or-nothing” simplification is removed. ~~SDs N_0 super-droplets~~ are initialized by randomly drawing radii from the assumed initial ~~distribution droplet size distribution, where N_0 is the initial number of real droplets in a cell.~~ Coalescence causes one of 15 the SDs to be discarded. ~~Time step~~ Unlike in the original method of Shima et al. (2009), ~~timestep~~ length is adapted at each ~~time step timestep~~ to ensure that none of the collision pairs has coalescence probability greater than 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 simulations, in which number of SDs ~~is constant remains constant (with rare exceptions)~~, is closer to 20 the ideas of Shima et al. (2009). We will refer to it as the “constant SD” simulations. In this type of simulations, the number of SDs is prescribed, ~~but they as N_{SD} and SDs~~ have different multiplicities. ~~To~~ Typically, N_{SD} is much smaller than N_0 . ~~We use a novel algorithm for initialising the radii and multiplicities of SDs. The aim of this algorithm is to~~ avoid large differences in the initial droplet size ~~distribution between realizations, SD sizes distributions between realizations.~~ Super-droplet radii are not completely randomly drawn from the assumed distribution as in the “one-to-one” simulations. Instead, the assumed 25 distribution is divided into N_{SD} bins and the radius of a single SD is randomly selected within each bin. The ~~bins have equal sizes on a logarithmic scale. Consider an initial droplet size distribution $n(\ln(r))$. We employ a notation in which we omit the division of radius by unit of length whenever the logarithm of radius is taken, i.e. $\ln(r)$ stands for $\ln(r/\mu\text{m})$. Concentration of droplets with radii in the range from r to $r + dr$ is $n|_{r, r+dr} = n(\ln(r)) d\ln(r)$. The~~ first step of the initialization is ~~finding to find~~ the largest and ~~the~~ smallest initial super-droplet radius, ~~r_{max} and r_{min}~~ r_{max} and r_{min} . They are found iteratively, starting 30 with ~~$r_{min} = 10^{-9}$ m and $r_{max} = 10^{-3}$~~ $r_{min} = 10^{-9}$ m and $r_{max} = 10^{-3}$ m. We require that they satisfy the condition

$$n(\ln(r_e)) \Delta l_r V \geq 1,\tag{2}$$

where r_e is either r_{max} or r_{min} , $n(\ln(r))$ is the initial droplet size distribution and $\Delta l_r = (\ln(r_{max}) - \ln(r_{min}))/N_{SD}r_{max}$ or r_{min} and $\Delta l_r = (\ln(r_{max}) - \ln(r_{min}))/N_{SD}$. In each iteration, if $r_{min} (r_{max} r_{min} (r_{max}))$ does not satisfy (2), it is increased (decreased) by 1%. Once r_{min} and r_{max} are found, radius of one SD is randomly selected within each bin of size Δl_r . Its multiplicity is given by $\xi = n(\ln(r))\Delta l_r V$. This N_{SD} super-droplets are created. Radius of the i -th SD is initialised by randomly selecting $\ln(r_i)$ in the range $(\ln(r_{min}) + (i-1)\Delta l_r, \ln(r_{min}) + i\Delta l_r]$. Initial multiplicity of the i -th SD is $\xi_i = \lfloor n(\ln(r_i))\Delta l_r V + 0.5 \rfloor$. Please note that increasing N_{SD} causes Δl_r to decrease and this in turn gives relatively large values of r_{min} and relatively small values of r_{max} . It means that this initialisation procedure does not represent well the tails of the distribution, especially for large N_{SD} . It also means that the “constant SD” initialisation with $N_{SD} = N_0$ is not equivalent to the “one-to-one” initialisation. Since the large tail is important for coalescence, we draw additional $\lfloor \int_{\ln(r_{max})}^{\infty} n(\ln(r)) d\ln(r) + 0.5 \rfloor$ add $\lfloor \int_{\ln(r_{max})}^{\infty} n(\ln(r)) d\ln(r) V + 0.5 \rfloor$ super-droplets with $\xi = 1$ to the cell. Their radii are selected by randomly drawing $\ln(r)$ from the distribution for radii greater than $r_{max} C n(\ln(r)) H(\ln(r) - \ln(r_{max}))$, where C is a normalizing constant and $H(x)$ is the Heaviside step function. This makes the actual number of SDs N 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 simulation, the time step simulations, the timestep 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 timestep length can make the coalescence probability exceed unity. If it does, it represent multiple collisions between is assumed that a pair of SDs (Shima et al., 2009). If multiplicity of a SD collides more than once during the timestep (Shima et al., 2009). Then, the procedure for calculating post-collisional parameters (Eq. 1) is applied $\tilde{\gamma} = \min(\gamma, \xi_i/\xi_j)$ times, where $\gamma \geq 1$ is the number of collisions between the i -th and the j -th SD and $\xi_i \geq \xi_j$. Such handling of multiple collisions can cause the expected number of collisions to be lower than the real one if $\gamma > \xi_i/\xi_j$. Another inconsistency is that rigorously, the probability of collision between SDs should change after each of the $\tilde{\gamma}$ collisions. For these reasons timestep length should be carefully selected so that multiple collisions do not happen often. If two SDs with identical ξ collide, multiplicity of one of them drops to zero, it. Then, the SD with $\xi = 0$ is used to split the SD with the largest ξ in the system into two. This is slightly different than in the Shima et al. (2009) algorithm, in which a SD with $\xi = 0$ is used to split the other SD that came out of the collision that caused the multiplicity to drop to zero. 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 Comparison of the “one-to-one” SDM with the master equation

The goal of this Section is to show that the “one-to-one” SDM is at the same level of precision as the master equation. To this end, we calculate average droplet size distribution and standard deviation of mass of the largest droplet from an ensemble of “one-to-one” simulations. As a reference, we use the results from Alfonso and Raga (2017), who used the master equation approach to study the sol-gel transition. 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 (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 to test if 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 $17 \text{ }\mu\text{m}$ and 10 droplets of radius $21.4 \text{ }\mu\text{m}$. Gravitational collision kernel is used with collision efficiencies from 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).

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~~ The average is calculated from an ensemble of $\Omega = 10^4$ realizations for each case. In simulations without linear sampling, all $N(N-1)/2$ collision pairs are considered and a constant ~~time-step timestep~~ $\Delta t = 0.1 \text{ s}$ ~~was is~~ used. Both ~~approaches give the same~~ approaches give similar results, what shows that the linear sampling technique does not affect the average number of collisions. In addition, the “one-to-one” simulations are compared with the master equation ~~approach~~ solved using the method of Alfonso and Raga (2017). Both approaches are generally in agreement, with some differences at the large end of the distribution. 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” 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. ~~One-to-one~~ This value is of interest, because 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, 2017). In Fig. 2, “one-to-one” simulations, with and without linear sampling, are compared with ~~SSA simulations~~ results of the master equation approach presented in Alfonso and Raga (2017). Please note that Alfonso and Raga (2017) obtained values of $\sigma(m_{\max})/\langle m_{\max} \rangle$ from an ensemble of SSA simulations rather than by solving the master equation, as was the case in Fig. 1. 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 of~~ 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.

The “one-to-one” 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 $n(m) = \frac{n_0}{\bar{m}} \exp(-m/\bar{m})$, where $n(m)dm$ is the number of droplets in mass range $(m, m+dm)$ in unit volume, $n_0 = 142 \text{ cm}^{-3}$ and \bar{m} is the mass of a droplet with radius $\bar{r} = 15 \text{ }\mu\text{m}$. This is the same distribution as in 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 10^5 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 cell 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

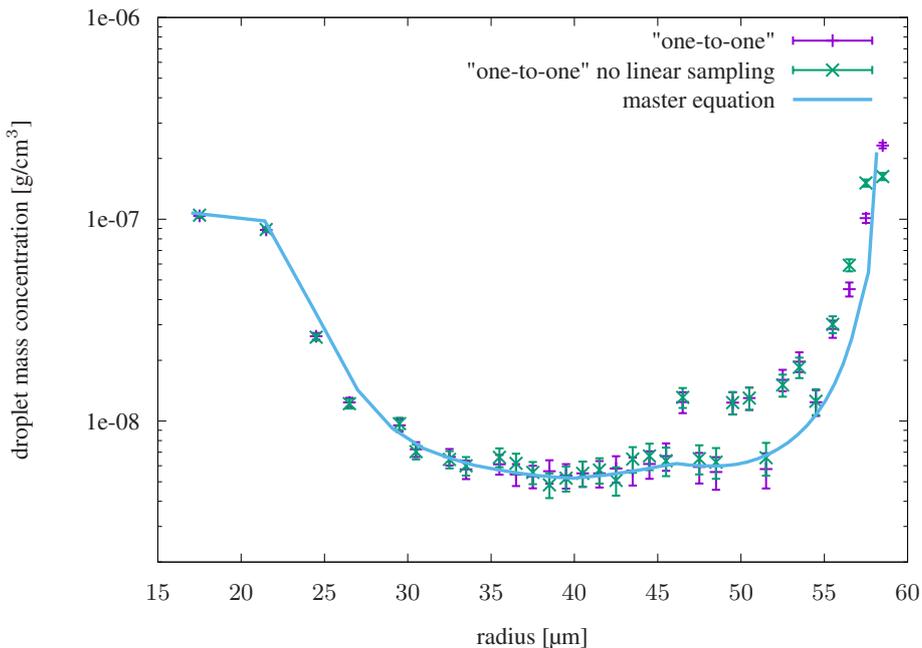


Figure 1. Mass of droplets per size bin at $t = 2500$ s. Bins are $1 \mu\text{m}$ wide. Points depict an averaged result of $\Omega = 10^4$ “one-to-one” simulations with and without linear sampling of collision pairs. Error bars show a 95% confidence interval. Line depicts a numerical solution of the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

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. ~~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, 2017). Intuitively, we would expect the time for most of the mass to accumulate in a single agglomerate to increase with increasing system cell size. This turns out to be true for systems cells with $N_0 > 10^3$. For system cell sizes $10^2 < N_0 < 10^3$ gelation time is approximately the same, around 300 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). Following their notation, by $t_{10\%}$ we denote time after which 10% of mass of cloud droplets is turned into droplets with $r > 40 \mu\text{m}$. 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 deter-

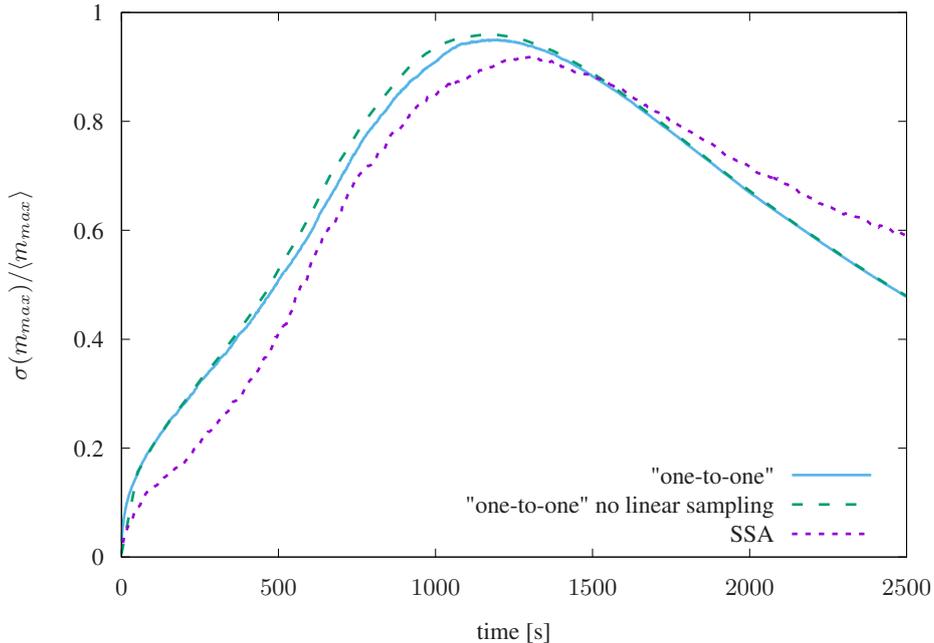


Figure 2. Relative standard deviation of mass of the largest droplet in the system. Details of the SDM simulations are given in the caption of Fig. 1. Size of the ensemble of SSA simulations is $\Omega = 10^3$. The SSA results are taken from Fig.7 in Alfonso and Raga (2017) (data courtesy of L. Alfonso).

mine how many SDs are needed in the “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 $t_{10\%}$ 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 the “constant SD” (rest of the points in the series) simulations are presented. For comparison, $t_{10\%}$ obtained by solving the Smoluchowski equation using 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 Bott simulation presented in this manuscript. Convergence tests were done for each case. The “one-to-one” results converge with increasing system-cell 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. In the “constant SD” simulations, error caused by using SDs with $\xi > 1$ weakly depends on the system-cell 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% 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.

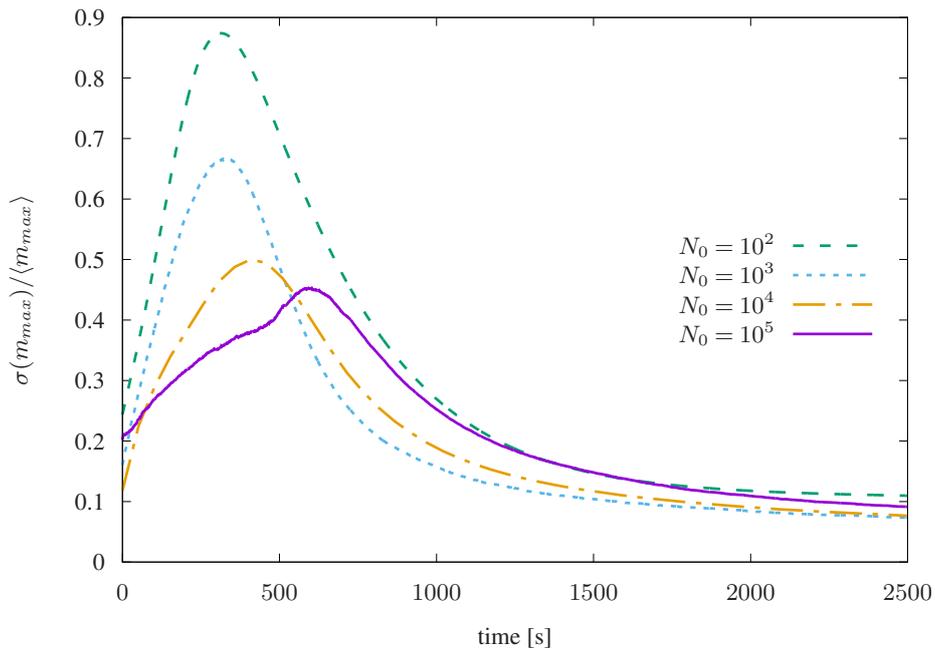


Figure 3. Relative standard deviation of mass of the largest droplet for different ~~system-cell~~ sizes. Estimated from ensembles of $\Omega = 10^4$ “one-to-one” simulations for each value of N_0 .

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” method, we plot the relative standard deviation of $t_{10\%}$ in Fig. 5. For reference, results of DNS from Onishi et al. (2015) are shown. Results from our “one-to-one” simulations are in good agreement with ~~them~~the DNS. Small discrepancies are probably caused by the fact that the DNS included turbulence of various strength for different N_0 . Results of the “one-to-one” simulations were fitted with the function $\alpha\sqrt{1/N_0}$, resulting in $\alpha = 6$. Figure 5 also presents fluctuations in 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 decreases correctly. Then, above some critical value of the N_0/N_{SD} ratio, fluctuations stop to decrease and remain constant independent of the ~~system-cell~~ size. This is a result of introducing unrealistic correlations between droplet sizes, which is a consequence of ~~the using~~ low number of ~~simulational-computational~~ particles (Bayewitz et al., 1974). To show that the linear sampling technique does not contribute to this effect, we plot result of a “constant SD” simulation without linear sampling for $N_{SD} = 32$, which is the same as for $N_{SD} = 32$ with linear sampling. 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 $t_{10\%}$ ~~using “constant SD” simulations, the~~ number of SDs

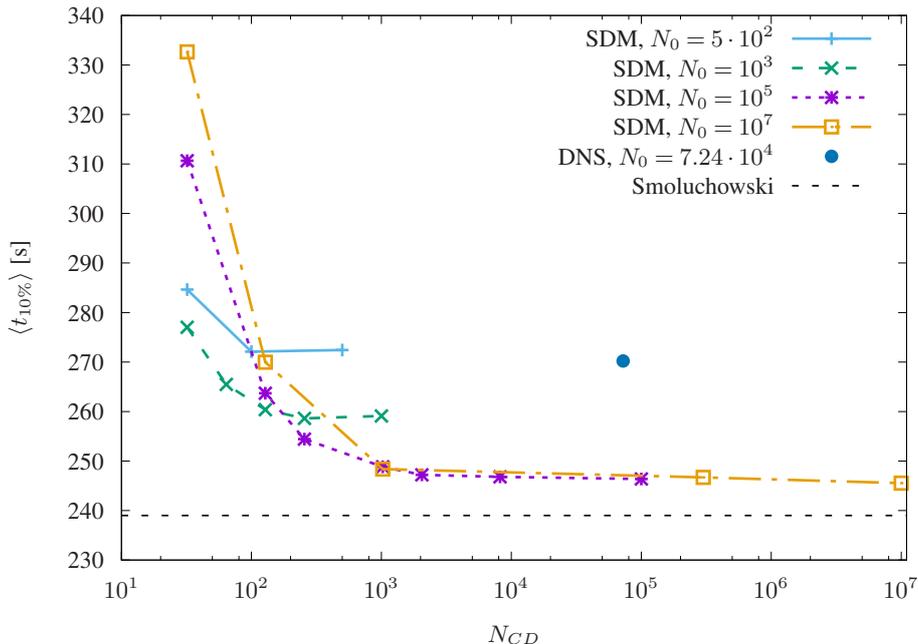


Figure 4. Mean $t_{10\%}$ for different system-cell sizes and different numbers of computational droplets N_{CD} . In SDM simulations, $N_{CD} = N_{SD}$ and in DNS, $N_{CD} = N_0$. The single DNS result is taken from Onishi et al. (2015) (the NoT-HI case therein). Ensemble sizes are $\Omega \geq 10^3$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

has to be $N_{SD} \geq \frac{1}{9}N_0$. Using so many SDs is not feasible in 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 ($V \rightarrow \infty$). We will try to determine minimal system-cell 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 ($r \geq 40 \mu\text{m}$) content to the total water content. 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 for two cases - with fast and with slow rain development. In both cases collision efficiencies for large droplets are taken from Hall (1980) and for small droplets from Davis (1972). In the-first-casesimulations with fast development of rain, we use the same initial distribution

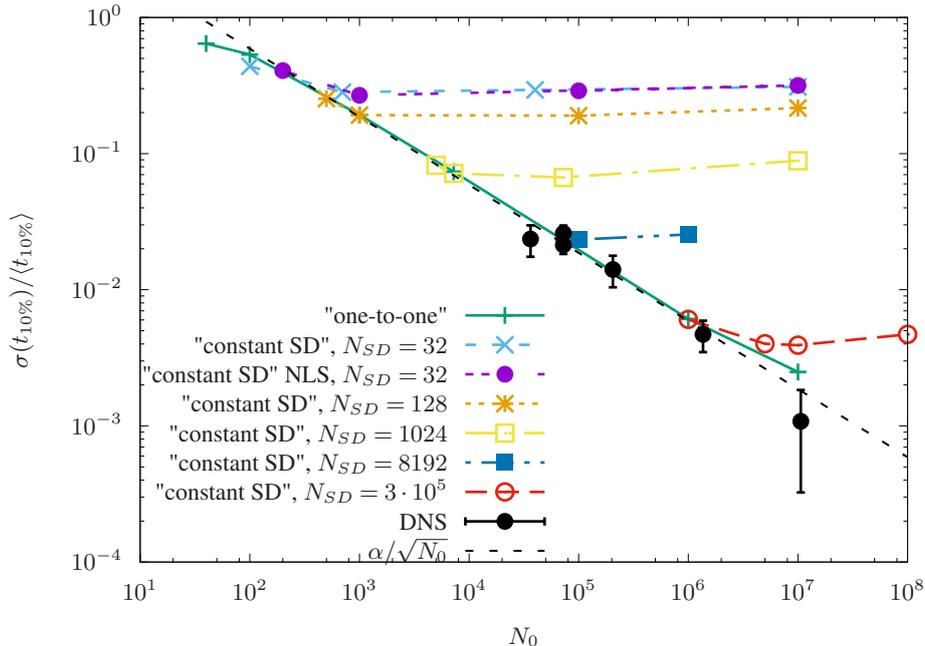


Figure 5. Relative standard deviation of $t_{10\%}$ against system-cell size. SDM results are based on samples of at least 10^3 realizations. DNS results are taken from Onishi et al. (2015). Where not shown, errorbars are smaller than plotted points. The value $\alpha = 6$ was obtained by curve fitting to the “one-to-one” results. The acronym NLS in the legend stands for “no linear sampling”.

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-cells with $N_0 \geq 10^4$. The Smoluchowski curve is slightly shifted left, probably due to the numerical diffusion of the Bott algorithm, as discussed in Sec. 4. In systems-cells smaller than $N_0 = 10^4$, rain develops slower than predicted by the Smoluchowski equation. Agreement of stochastic coalescence in large systems-cells with the Smoluchowski equation for a similar initial distribution was shown in-using the SSA by Seesselberg et al. (1996). Onishi et al. (2015) present figures similar to Fig. 7, but obtained from DNS runs for $N_0 = 7.24 \cdot 10^4$ (Fig. 1(b) therein). They also show good agreement between model-results-DNS and the Smoluchowski equation with the kernel of Long (1974), at least up to $t = 330$ s. If the Hall kernel is used in the Smoluchowski equation, autoconversion is quicker than in the DNS, as discussed in Sec. 4.

The second case is well-Next, we validate the Smoluchowski equation in a setup with slow rain development. This time the initial droplet size distribution is below the size gap, i.e. the range of radii for which both collisional and condensational growths are slow. We use $\bar{r} = 9.3 \mu\text{m}$ and $n_0 = 297 \text{ cm}^{-3}$ as in Wang et al. (2006). In addition, we cut the distribution to 0 at $r = 20$ for $r > 20 \mu\text{m}$. This cutoff is used in the SDM modelling as well as when solving the Smoluchowski equation. That way we get rid of the occasional very large SDs present at $t = 0$ in some realizations 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. Again, we see

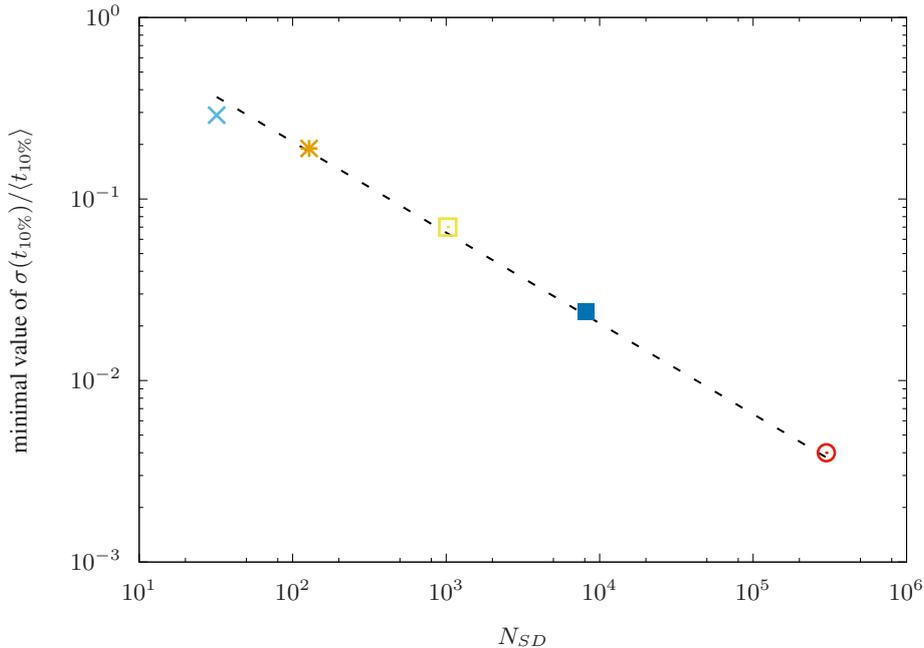


Figure 6. Minimal points depict the minimal, limiting value of the relative standard deviation of $t_{10\%}$ for a given number of super-droplets (squares) in “constant SD” simulations. For each value of N_{SD} , the minimal value of $\sigma(t_{10\%})/(t_{10\%})$ is calculated as an average of the respective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

convergence of the “one-to-one” simulations to the Smoluchowski result, but in this case the system-cell has to be larger ($N_0 \geq 10^7$) for the 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-cells rain appears later than in larger systems-cells. On the other hand, the rain formation rate (the slope of the curves in Fig. 8) in smaller systems-cells starts to decrease at higher values of θ than in larger systems-cells. In consequence, smaller systems-cells 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 rain formation rate is associated with the decrease in the concentration of rain drops n_r , plotted in Fig. 9. Number of rain drops decreases due to collisions between drops from this category. A single drop that is produced in such collision is less efficient at scavenging cloud droplets than the two pre-collision 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 immediately 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

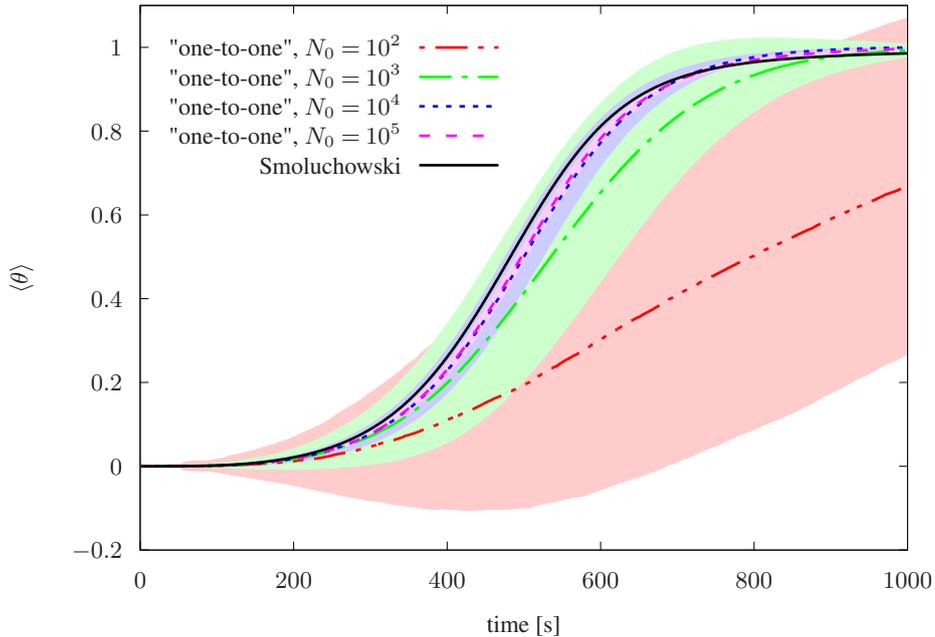


Figure 7. Rain content ratio θ for different ~~system-cell~~ sizes averaged over ensembles of $\Omega = 10^3$ simulations. Shaded regions show one standard deviation interval.

equation, may result in ~~underestimating the amount~~ an artificial decrease in the number of rain drops and an underestimation of mass 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. 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 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). They could also lead to the deviation from the $\sim 1/\sqrt{N_0}$ scaling seen shown in Fig. 5. Fluctuations in cells with $N_0 = 10^7$ are greater than predicted using this scaling. We also observe that ~~although the amount of rain water depends strongly for $t < 3000$ s, the number of rain drops does not~~ depend on cell size ~~the number of rain drops does not~~ (c.f. Fig. 9), but the mass of rain water does (c.f. Fig. 8). In larger cells rain drops acquire larger sizes through collisions with cloud droplets, but rate of autoconversion of cloud droplets to rain drops is not affected much by cell size.

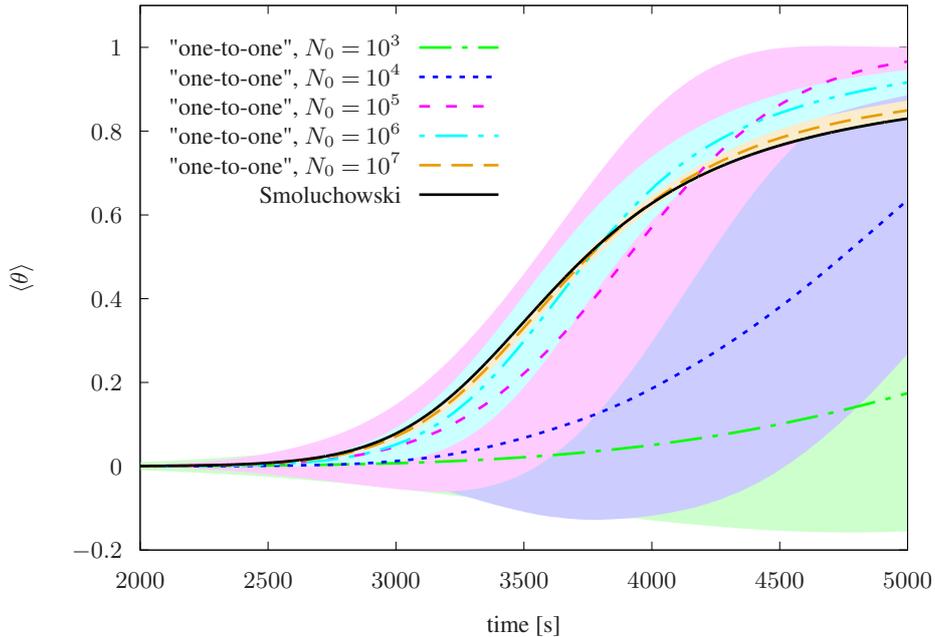


Figure 8. As in Fig. 7, but for an initial distribution with $\bar{r} = 9.3 \mu\text{m}$, $n_0 = 297 \text{ cm}^{-3}$ and a cutoff at $r = 20 \mu\text{m}$. The ensemble-ensemble size is $\Omega = 10^8/N_0$.

Table 1. Average $\langle t_{40} \rangle_\gamma$ and standard deviation $\sigma(t_{40})_\gamma$ and sample size of time (in seconds) for the lucky realizations to produce a single rain-drop with $r > 40 \mu\text{m}$. γ is fraction of the fastest realizations out of all Ω realizations that were used to compute the respective $\langle t_{40} \rangle_\gamma$ and $\sigma(t_{40})_\gamma$. The sub-ensemble size $\gamma\Omega$ is shown to give an idea about precision of estimation of the respective statistics.

	$\gamma = 10^{-4}$			$\gamma = 10^{-3}$			$\gamma = 10^{-2}$			$\gamma = 10^{-1}$			$\gamma = 1$		
N_0	$\langle t_{40} \rangle_\gamma$	$\sigma(t_{40})_\gamma$	$\gamma\Omega$												
10^2	2052	212	10	2930	356	10	4053	517	10^2	6365	1158	10^3	14777	6099	10^3
10^3	1366	120	10^2	1762	170	10^3	2400	267	10^4	3440	505	10^5	6500	1700	10^6
10^4	1089	173	3	1336	103	10	1717	176	10^2	2354	276	10^3	3912	764	10^4
10^5	946	33	2	1090	60	20	1334	85	200	1721	169	2000	2552	415	10^4
10^6							1038	165	2	1301	176	20	1831	277	10^2

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 argued to be responsible for droplet spectra broadening and rain forming quicker than predicted by the Smoluchowski equation.

5 Luck is supposed to be especially important during crossing of the size gap, when collisions happen rarely (Robertson, 1974;

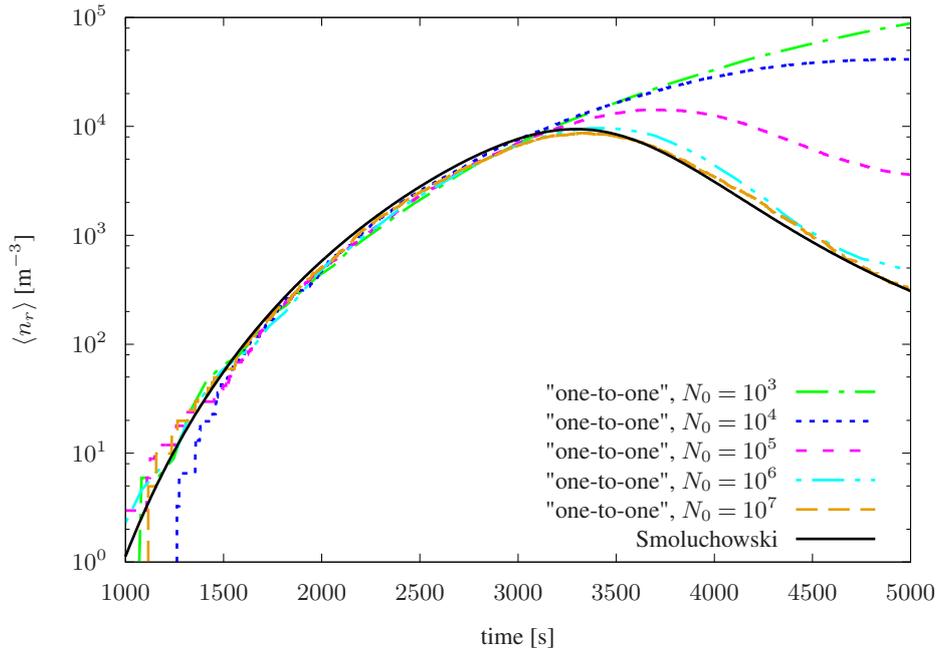


Figure 9. Mean concentration of rain drops from the same simulations as in Fig. 8.

Kostinski and Shaw, 2005). A single droplet that would cross the size gap through lucky collisions could then initiate a cascade of collisions. Theoretical estimation of the “luck factor” was presented in 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 until the first droplet grows to $r = 40$ μm . We perform simulations for the same initial distribution ensembles of simulations for different cell sizes N_0 . Size of an ensemble is denoted by Ω . The initial distribution is the same as in the second case in Sec. 5. The mean radius is $\bar{r} = 9.3$ μm , well below the size gap. From an ensemble of Ω realizations The liquid water content is $1\text{g}/\text{cm}^3$ and the concentration is 297 cm^{-3} , so the smallest cell that has enough water to produce a droplet with $r = 40\text{ }\mu\text{m}$ is $N_0 \approx 80$. Therefore the smallest cell size we consider is $N_0 = 10^2$. For each value of N_0 , 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 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 statistics were calculated. The results for different cell sizes are shown in Tab. 1.

There is a large variability in $\langle t_{40} \rangle_\gamma$ 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 $\langle t_{40} \rangle_\gamma$ is approximately the same along the diagonals of Tab. 1. For example, a 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 out of ten 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 $\langle t_{40} \rangle_\gamma$ 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~~ To depict that size of a coalescence cell does not affect formation of first rain drops, it is helpful to think

5 ~~about the test presented in this Section as a simulation of a large system initially containing $N_{\text{tot}} = N_0/\gamma$ droplets divided into coalescence cells of size N_0 . We are interested in the mean time $\langle t_{40} \rangle$ for the first droplet out of all N_{tot} droplets to grow to $r = 40 \mu\text{m}$. In Fig. 10 we plot $\langle t_{40} \rangle$ against N_{tot} for different sizes of coalescence cell. It is seen that $\langle t_{40} \rangle$ does not depend on N_0 . Even using a coalescence cell with $N_0 = 10^2$ can produce has $r \approx 43 \mu\text{m}$, close to the $40 \mu\text{m}$ rain threshold, in which there is barely enough water to produce a drop with $r = 40 \mu\text{m}$, does not change the results.~~

10 ~~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 Next we calculate the “luck factor”, i.e. how much faster the luckiest droplets grow to $r = 40 \mu\text{m}$ compared to the average droplets. To calculate it we use the data for $N_0 = 10^2$, because cells of this size can produce only a single droplet with $r = 40 \mu\text{m}$. Larger~~

15 ~~cells behave like an ensemble of cells with $N_0 = 10^2$ as far as t_{40} is concerned, larger cells behave exactly like an ensemble of cells of this size, so calculating the “luck factor” using t_{40} from larger cells would tell us how much faster the luckiest ensemble of droplets produces a rain droplet, compared to an average ensemble - a quantity that we are not interested in. We find $\langle t_{40} \rangle_{10^{-3}} / \langle t_{40} \rangle_1 \approx 3.7$ and $\langle t_{40} \rangle_{10^{-5}} / \langle t_{40} \rangle_1 \approx 6$, $\langle t_{40} \rangle_1 / \langle t_{40} \rangle_{10^{-3}} \approx 5$ and $\langle t_{40} \rangle_1 / \langle t_{40} \rangle_{10^{-5}} \approx 11$. The value of $\langle t_{40} \rangle_{10^{-5}}$ was estimated at ~~1090–1366~~ s based on values along the diagonal for larger γ and larger N_0 . These ratios are lower than~~

20 ~~given in Kostinski and Shaw (2005), showing that their theoretical analysis slightly overestimates the “luckiness” in droplet growth. 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. These values are in good agreement with our observations.~~

7 Size of a well-mixed coalescence cell

In the previous Sections we have seen that size of the coalescence cell has a profound impact on the evolution of the system.

25 In this Section we estimate the size of a cell that can be assumed to be well-mixed. All methods in which 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 ~~$\tau_{\text{mix}} \ll \tau_{\text{coal}}$, where τ_{coal} and $\tau_{\text{mix}} \ll \tau_{\text{coal}}$, 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

30 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.

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

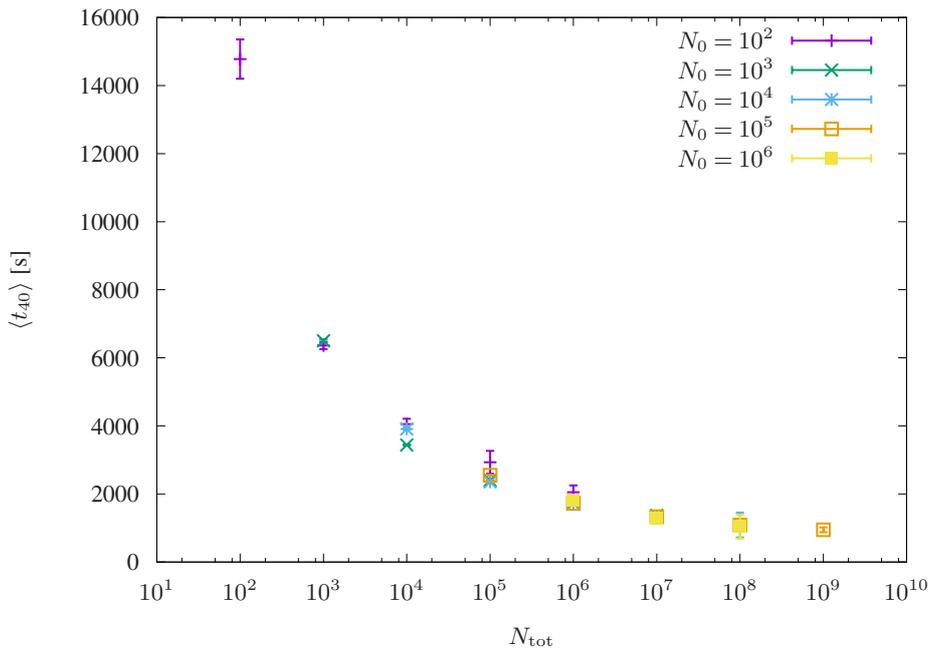


Figure 10. Mean time until a system of N_{tot} droplets produces first droplet with $r = 40\mu\text{m}$. The system is divided into coalescence cells of size N_0 . The figure is based on the results of “one-to-one” simulations given in Tab. 1.

a field of smaller droplets. Using geometric coalescence kernel with efficiency E , the mean time between collisions is $\tau_{coat} = (E\pi(r_l + r_s)^2 v_r n_s)^{-1} \tau_{coal} = (E\pi(r_l + r_s)^2 v_r n_s)^{-1}$, 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 $r_l = 100\mu\text{m}$, $r_s = 10\mu\text{m}$, $v_r = 70\text{cm/s}$, $E = 1$ and $n_s = 100\text{cm}^{-3}$ we get $\tau_{coat} \approx 0.4\text{s}$, $\tau_{coal} \approx 0.4\text{s}$.

- 5 Droplets in the cell can be mixed through turbulence. Turbulence acts similarly to diffusion and its characteristic time for mixing is $\tau_{mix}^t = (V^{(2/3)}/\varepsilon)^{(1/3)} \tau_{mix}^t = (V^{(2/3)}/\varepsilon)^{(1/3)}$, 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 $10\text{cm}^2/\text{s}^3$ for stratocumulus clouds to $10^3\text{cm}^2/\text{s}^3$ for cumulonimbus clouds (Malinowski et al., 2013; Grabowski and Wang, 2013). Let us assume that $\tau_{mix}^t \ll \tau_{coat} \ll \tau_{coal}$ is satisfied if $\tau_{mix}^t = 0.1\tau_{coat}$, $\tau_{mix}^t = 0.1\tau_{coal}$. Even in the most turbulent clouds, this means that
- 10 the coalescence cell has to be very small $V \approx 1.5 \cdot 10^{-2}\text{cm}^3$. On average, this volume would contain around one droplet, depending on concentration of droplets. For such small coalescence volumes, the Smoluchowski is not valid and SDM would be very cumbersome, because extremely short **time steps timesteps** would be required. To use larger cells, we need to choose some less strict value of characteristic time of coalescence. Some larger cell size, that would be approximately well-mixed, could be found phenomenologically through fine-grid simulations including droplet motion. One example of such reference
- 15 simulations are DNS runs from Onishi et al. (2015) discussed in Sec. 5. They prove that in the case with $\bar{r} = 15\mu\text{m}$, 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
5 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
10 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 ($\bar{r} = 9.3 \mu\text{m}$) and with large ($\bar{r} = 15 \mu\text{m}$) droplets. They result in slow and fast rain formation, respectively. Dependence of the system behavior
15 on 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 values. If the coalescence cell is small, sizes of droplets are strongly correlated and depletion of cloud water plays an important role. In real clouds, these two effects are probably not manifested, because collector drop sedimentation acts against them. In relatively large cells, rain drops 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
20 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 ~~unrealistically~~ unrealistically large well-mixed 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 ~~equation~~ equation, produce too little rain.

25 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 the increase in the rate of collisions between rain drops is not important. The mean behavior of the system converges to the Smoluchowski equation results with increasing ~~system-cell~~ system-cell size. Good agreement with it is found for ~~systems-cells~~ systems-cells 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 the additional collisions is relatively more
30 important. 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 3.75 times faster than average and the luckiest out of a hundred thousand - 6.11 times faster. These values are ~~slightly smaller than the~~ are in good agreement with the analytical estimation of Kostinski and Shaw (2005).

We estimate a well-mixed (with respect to coalescence) volume in the most turbulent clouds to be only $1.5 \cdot 10^{-2} \text{ cm}^3$. 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 $N_0 = 10^4$ droplets also gives the same results. Therefore it can be assumed that such volume is approximately well-mixed in the case of fast coalescence. In the slow-coalescence case, the well-mixed volume needs to be larger than in the fast-coalescence case for the Smoluchowski equation to be valid. Size of an approximately well-mixed cell for this case can be determined using DNS with initially small droplets. ~~Cells~~ Volume of cells used in LES ~~are~~ is typically ten orders of magnitude larger than a well-mixed volume. ~~They~~ The LES cells 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.

9 Code availability

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

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 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. (2017). The impact of fluctuations and correlations in droplet growth by collision–coalescence revisited – part 5 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.
- 10 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*, 15 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.
- 20 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 25 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.
- 30 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.
- Long, A. B. (1974). Solutions to the droplet collection equation for polynomial kernels. *Journal of the Atmospheric Sciences*, 31(4):1040–1052.
- 35 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.
- Seesselberg, 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.
- 10 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 Vorträge über Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen. *Zeitschrift für Physik*, 17:557–585.
- 15 Sölch, I. and Köhler, B. (2010). A large-eddy model for cirrus clouds with explicit aerosol and ice microphysics and Lagrangian ice particle tracking. *Quarterly Journal of the Royal Meteorological Society*, 136(653):2074–2093.
- 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.
- 20 Unterstrasser, S., Hoffmann, F., and Lerch, M. (2017). Collection/aggregation algorithms in Lagrangian cloud microphysical models: rigorous evaluation in box model simulations. *Geoscientific Model Development*, 10(4):1521.
- 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.