

Reply to Reviewer:

First, we would like to thank the anonymous referee for his/her comments that surely will improve the quality of our paper. Our revised version will include many of his/her suggestions.

**1) Specific Comments:**

In the revised version of our paper, we will expand the list of previous papers that addressed the lucky droplet model for the collision coalescence process (by including some papers suggested by the reviewer), and also will explicitly explain the main differences between previous studies and present work, both in the introduction and in the conclusions.

**On the main differences between previous studies and present work:**

Previous efforts on this direction were mainly focused on finding the distribution of times for  $N$  collisions (Telford, 1955; Kostinski and Shaw, 2005; Wilkinson; 2016), while we were concentrated on studying the “lucky droplet” size distribution to determine whether or not the runaway growth process due to collision-coalescence has started.

For example, Kostinski and Shaw (2005) present a distribution of the time to produce drizzle by calculating the convolution of the exponentially distributed times between collisions. They found the distribution of:

$$T_{N_c} = \sum_{i=1}^{N_c} t_i \quad (S1)$$

with  $N_c$  fixed, where the  $t_i$  are the times between droplet collisions and  $N_c$  the number of collisions, which have an exponential distribution.

Wilkinson (2016) found the probability density for the time  $T_{N_c} = \sum_{i=1}^{N_c} t_i$  being a small fraction of its average value (far from the mean value). As the precipitation occurs on a time scale that is smaller than the typical scale for one collision, the problem was solved by applying large deviation theory. **More details on large deviation theory (LDT) can be found in this document in the reply to question 20.**

**References:**

Kostinski, A. B., & Shaw, R. A. (2005). Fluctuations and luck in droplet growth by coalescence. *Bulletin of the American Meteorological Society*, 86(2), 235-244.  
Telford, J. W. (1955). A new aspect of coalescence theory. *Journal of Meteorology*, 12(5), 436-444.  
Wilkinson M. 2016. Large deviation analysis of rapid onset of rain showers. *Phys. Rev.*

**Reviewer:**

2. *Many papers have studied the collision-coalescence problem, which should be also addressed in the introduction. A good summary is given by Grabowski and Wang (2013) [5]. The work (summarized in Grabowski and Wang (2013) [5]) by the group of Wang should be addressed. Several stochastic models by Pinsky et al (2004, 2007, 2008) [8]-[10], Mehlig et al (2007) [6], and Wilkinson et al (2006) [7] should be cited. Recent numerical work by Onishi and Seifert (2016) [11], Li et al (2017) [12], Li et al (2018) [13], and Chen et al. (2018) [14].*

**Reply:**

As suggested by the reviewer, a list of papers that addressed the collision-coalescence problem will be added.

**Reviewer:**

*I would suggest the authors compare the Monte-Carlo method used in Shima et al. 2009 [15], Li et al (2017) [12] and Li et al (2018) [13].*

**Reply:**

**The Monte Carlo algorithm (Difference between the algorithm of Gillespie (1976) and the algorithm of Shima (2009)):**

In our study we use the stochastic simulation algorithm (SSA) developed by Gillespie (1976) for chemical reactions which rigorously account for fluctuations and correlations in a coalescing system. This algorithm was reformulated to simulate the kinetic behavior of aggregating systems by Laurenzi and Diamond (2002), by defining species as a type of aggregate with a specific size and composition. In our case, species represent droplets of different sizes.

The main difference between the Gillespie's SSA and the Monte Carlo method used in Shima et al. (2009), is that the SSA involved the collision of only two physical particles (droplets in our case) per MC cycle, while in the Super Droplet (SD) method developed by Shima et al. (2009) and other algorithms based on the simulation particles (SIP) approach (Li et al, 2017), in each MC cycle collide super-droplets, that represents a multiple number of droplets with the same attributes (radius  $r$  or mass in the simplest case) and position.

For Gillespie's SSA the number of collisions ( $C_T$ ) during a time interval  $\Delta t$  can be estimated from the expression:

$$C_T = \frac{\Delta t \sum_{j=1}^{N_s} \sum_{i=1}^{N_s} K(i, j) N_i N_j}{V} \quad (\text{S2})$$

In (1)  $K(i, j)$  is the collection kernel,  $V$  is the coalescence cell volume and  $N_i$  are the number of particles in species with index  $i$  (particles of the same radius).

As the number of collisions  $C_T$  (see Eq. 1) will increase quadratically with the initial number of particles (Gillespie, 1975), we can conclude that the application of the SSA in systems involving a large number of particles, and with only two physical particles colliding per MC cycle is highly impractical. For example, in a three dimensional cloud model the typical coalescence cell has a volume of  $10^9 \text{ cm}^3$  and considering a droplet concentration at cloud base typical of maritime clouds ( $10^2 \text{ cm}^{-3}$ ), then the number of droplets will be about  $10^{11}$ . Then, in this case the Gillespie's SSA is not a suitable option due to the huge number of collisions in large volumes, and the high cost in computation.

The super-droplet method of Shima et al. (2009) was design to overcome this problem. The total collision rate in a time interval for this method can be calculated from the expression:

$$C_{T-SD} = \frac{\Delta t \sum_{j=1}^{N_{SD}} \sum_{i=1}^{N_{SD}} K(i, j) \max(\xi_i, \xi_j)}{V} \quad (\text{S3})$$

where (2)  $\xi_i$  is the super-droplet's multiplicity (number of physical droplets of the same radius) and  $N_{SD}$  the number of super-droplets. From comparison of (S1) and (S2) it can be concluded that the number of super-droplet's collisions in a time interval it's much smaller. In the original paper Shima et al. (2009) compared the results obtained with Monte Carlo simulations with the SD method with numerical solutions of the kinetic collection equation and confirmed that the SD method reproduces the solution of the KCE if the number of super droplets  $N_{SD}$  is sufficiently large ( about  $2^{17}$ ). However, Unterstrasser et al. (2017) show that convergence is possible with a number of super droplets in the order of  $10^2$ . Another simplification of the SD method is that instead of considering  $N(N-1)/2$  collision pairs only  $[N/2]$  non overlapping randomly selected pairs are considered.

As was stated before, the SSA of Gillespie (1975, 1976) rigorously account for fluctuations and correlations in a coalescing system, and the temporal evolution of mean values at each droplet size can be obtained by averaging over many runs. However, in order to obtain accurate solutions at the large end of the distribution, a large number of realizations is required. The alternative is the master equation (Bayewitz et al, 1974; Alfonso, 2015; Alfonso and Raga, 2017), which also accounts for fluctuations and correlations, and can serve as a reliable benchmark for different Monte Carlo methods.

Dziekan and Pawlowska (2017) performed "one to one" simulations (in that case the multiplicity is  $\xi_i=1$ , and  $SD$  is equivalent to a physical droplet) with the SD method of Shima (2009), compared the solutions with the master equation (Alfonso and Raga, 2017) and found that both approaches are generally in agreement, only with some differences at the large end of the distribution. Simulations results by Unterstrasser (2018) also show a good correspondence with the master equation even at the large end of the droplet size distribution. Gillespie's (1975, 1975) SSA works perfect for our purposes because it rigorously account for fluctuations and correlations that are inherent to a finite system. **Due to the finiteness of the systems**, our simulations are performed for small volumes with small number of droplets (in the range  $50\text{-}300 \text{ cm}^{-3}$ ).

**References:**

- Alfonso, L.: An algorithm for the numerical solution of the multivariate master equation for stochastic coalescence. *Atmospheric Chemistry and Physics*, vol. 15, no 21, p. 12315-12326, 2015.
- Alfonso, L. and Raga, G.B.: The impact of fluctuations and correlations in droplet growth by collision-coalescence revisited. Part I: Numerical calculation of post-gel droplet size distribution, *Atmos Chem. Phys.*, 17, 6895–6905, 2017.
- Bayewitz, M.H., Yerushalmi, J., Katz, S., and Shinnar, R.: The extent of correlations in a stochastic coalescence process, *J. Atmos. Sci.*, 31, 1604-1614, 1974.
- Dziekan, P., & Pawlowska, H. (2017). Stochastic coalescence in Lagrangian cloud microphysics. *Atmospheric Chemistry and Physics*, 17(22), 13509-13520.
- Gillespie, D.T.: An Exact Method for Numerically Simulating the Stochastic Coalescence Process in a Cloud, *J. Atmos. Sci.* 32, 1977-1989, 1975.
- Gillespie, D.T.: An Exact Method for Numerically Simulating the Stochastic Coalescence Process in a Cloud, *J. Atmos. Sci.* 32, 1977-1989, 1975.
- Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22(4), 403-434.
- Laurenzi, I. J., Bartels, J. D., and Diamond, S. L.: A general algorithm for exact simulation of multicomponent aggregation processes. *Journal of Computational Physics*, 177(2), 418-449, 2002.
- Li XY, Brandenburg A, Haugen NEL, Svensson G. 2017. Eulerian and Lagrangian approaches to multidimensional condensation and collection. *J. Adv. Modeling Earth Systems* 9(2): 1116–1137.
- Shima, S. I., Kusano, K., Kawano, A., Sugiyama, T., & 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.
- Unterstrasser, S., Hoffmann, F., & Lerch, M. (2017). Collection/aggregation algorithms in Lagrangian cloud microphysical models: rigorous evaluation in box model simulations. *Geoscientific Model Development*, 10(4), 1521.

**Reviewer:**

3. For the fitted distribution in Fig.1, 5, and 7, could the authors have more samples to get better statistics?

**Reply:**

The number of realizations in our Monte Carlo algorithm, will be the sample size in the application of the Block Maxima approach (Figs 1 and 5 of the paper). That's why do not increase the number of realizations when generating synthetic data, taking into account that the number of realizations in the simulations (1000) must be close to the average number of blocks for which the largest droplet maxima can be fitted to the combined distribution (Eq. 6 of the paper).

On the other hand (analyzing the problem in terms of the accuracy needed for calculating the average values), the simulations were performed for 1000 realizations that is sufficiently to obtain the desired accuracy for the expected values:

$$\langle N \rangle = \frac{\sum_{i=1}^{N_r} N_i}{N_r} \quad (\text{S4})$$

The errors of the stochastic procedure can be calculated following Gillespie (1975) from the expression:

$$\sigma(N(t)) = \left\{ \frac{1}{K} \sum_{i=1}^K [N^i(t)]^2 - \left[ \frac{1}{K} \sum_{i=1}^K N^i(t) \right]^2 \right\} = \langle N^2 \rangle - \langle N \rangle^2 \quad (\text{S5})$$

Where  $\langle N \rangle$  the ensemble average and  $N_i$  is the droplet concentration for each realization. The ensemble average will be estimated with the desired accuracy if the condition

$$\sigma(N(t)) / \langle N(t) \rangle \ll 1 \quad (\text{S6})$$

fulfilled. The errors of the procedure can be checked in Fig. 5 of Alfonso et al. (2013), demonstrating that the Monte Carlo averages are calculated with the desired accuracy. For Fig. 7, the sample size (500) was set equal to the average number of blocks for which the largest droplet maxima can be fitted to the mixture of distributions.

#### References:

- Alfonso, L., Raga, G. B., and Baumgardner, D.: The validity of the kinetic collection equation revisited–Part 3: Sol–gel transition under turbulent conditions. *Atmospheric Chemistry and Physics*, vol. 13, no 2, p. 521-529, 2013.
- Gillespie, D.T.: An Exact Method for Numerically Simulating the Stochastic Coalescence Process in a Cloud, *J. Atmos. Sci.* 32, 1977-1989, 1975.

#### Reviewer:

4. A question related to question 3.: on Line 239, the authors used 200 droplets of 10um, and 50 droplets of 12.6 um for the Monte Carlo simulation. Is it statistically convergent? Can the authors provide a statistically convergent study (similar to the one in Li et al (2017) [12])?

#### Reply:

The idea was to perform simulations for small systems (with a small number of particles) for which fluctuations and correlations are relevant. That’s why the number of droplets per cm<sup>3</sup> use in the simulations are small, and of the same order of the droplet concentrations obtained from observations (which fluctuate between 0 and 392, with an average of 146). **This point will be clarified in the revised version of the paper.**

#### Reviewer:

5. L65: please provide reference for the use of “gel formation” in “percolation theory” and “nuclear physics” respectively.

#### Reply:

The corresponding references will be added in the revised version:

- Botet, R., Płoszajczak, M., Chbihi, A., Borderie, B., Durand, D., and Frankland, J.: Universal fluctuations in heavy-ion collisions in the Fermi energy domain. *Physical Review Letters*, 86(16), 3514, 2001
- Botet, R. and Płoszajczak, M.: Exact order-parameter distribution for critical mean-field percolation and critical aggregation. *Physical Review Letters*, 95(18), 185702, 2005.
- Gruyer, D., Frankland, J. D., Botet, R., Płoszajczak, M., Bonnet, E., Chbihi, A., ... and Guinet, D.: Nuclear multifragmentation time scale and fluctuations of the largest fragment size. *Physical review letters*, 110(17), 172701, 2013.

**Reviewer:**

6. L80: "average number of droplets". Do you mean "droplet (particle) number density" ? I would suggest the author use the commonly accepted terminology in both cloud physics and statistical mechanics for readability.

**Reply:**

As was stated by Gillespie (1972), the definition of  $N(i, t)$  in the kinetic collection equation (Eq. 1 of the paper), vary from author to author, but usually is taken to be **the average concentration of cloud droplets of mass  $i$  at time  $t$** . In the revised version we will clarify this point in order to avoid confusions, as suggested.

**References:**

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

**Reviewer:**

7. L81: I don't quite understand "the time rate of change of...". Could you please rephrase the sentence for readability?

**Reply:**

We will rewrite the sentence in the revised version as suggested.

**Reviewer:**

8. Eq.3, where is " $\tau$ " defined?

**Reply:**

There is a mistake in Eq.3, we should have written  $t$  instead of  $\tau$ .

**Reviewer:**

9. I don't understand how Eq.4 is obtained. What is  $T_{gel}$ ? What is the physics of this time scale?

**Reply:**

$T_{gel}$  is defined as the time when the second moment  $M_2(t) = \frac{M_2(t_0)}{1 - CM_2(t_0)t}$  becomes infinite,

then  $1 - CM_2(t_0)t = 0$ , and  $T_{gel} = [CM_2(t_0)]^{-1}$ . The equation for  $M_2(t)$  (moment of order 2) can be obtained from the general equation for moment evolution that was obtained by Drake (1972) from the continuous form of the KCE. It has the form:

$$\frac{dM_n(t)}{dt} = \frac{1}{2} \int_0^\infty \int_0^\infty [(x+y)^n - x^n - y^n] K(x,y) N(x,t) N(y,t) dx dy \quad (S7)$$

In (7),  $K(x,y)$  is the collection kernel  $N(x,t)$  is the average droplet concentration. If we consider the product kernel  $K(x,y) = C(xy)$  in equation (7), then, the equation for the second moment is:

$$\frac{dM_2(t)}{dt} = C[M_2(t)]^2 \quad (S8)$$

So that  $M_2(t) = \frac{M_2(t_0)}{1 - CM_2(t_0)t}$

After  $T_{gel}$ , the total, a runaway droplet forms, and the kinetic collection equation is no longer valid, since the assumption of a continuous distribution breaks down. There is in essence a phase transition in the system from a continuous distribution to a continuous distribution **plus** a runaway droplet.

**References:**

Drake, R.L.: The scalar transport equation of coalescence theory: Moments and kernels, J. Atmos. Sci., 29, 537-547, 1972.

**Reviewer:**

11. L110: please provide reference after “experimentally”.

**Reply:**

The corresponding reference was added:

Lushnikov, A. A., Negin, A. E., & Pakhomov, A. V. (1990). Experimental observation of the aerosol-aerogel transition. *Chemical physics letters*, 175(1-2), 138-142.

**Reviewer:**

12. L111: please provide reference after “percolation”.

**Reply:**

The corresponding reference will be added added:

Kolb, M., & Axelos, M. A. (1990). Gelation Transition versus Percolation Theory. In *Correlations and Connectivity* (pp. 255-261). Springer, Dordrecht.

**Reviewer:**

13. Eq. 5a and 5b, please compare them with Kostinski and Shaw (2005) [2] and Wilkinson (2016) [3].

**Reply:**

Our approach is different from that of the mentioned authors. The Kolmogorov distribution (5a, 5b) and the mixture of a Gaussian and a Gumbel (6) are the distributions of the largest cluster (droplet) mass at critical point (Eqs, 5a, 5b) and in the pseudocritical region (Eq. 6) respectively. Kostinski and Shaw (2005) and Wilkinson (2016) were interested in the distribution of **times for  $N$  collisions**.

**Reviewer:**

14. L133: “We must emphasize that phase transitions cannot take place in a finite system. For this type of systems, the notion of pseudo-critical region is introduced”. Please provide more physical explanation and references for the statement and “pseudo-critical region”.

**Reply:**

In theory of critical phenomena, a phase transition is defined as a singularity in the free-energy or any thermodynamic property of a system, which is proportional to the logarithm of the sum of exponentials. For finite-sized systems, the free energy is proportional to the logarithm of a finite number of exponentials, which are always positive. Then, such singularities are only possible within infinite systems by taking the thermodynamic limit  $N \rightarrow \infty$  (Bhattacharjee, 2001). For example, for the the Smoluchowski model (which is obtained in the thermodynamic limit), there is a disordered phase (before the sol gel transition), and ordered phase (after the sol gel transition), and  $T_{gel} = [CM_2(t_0)]^{-1}$  is the critical point for the infinite system. Then, a phase transition cannot take place in a finite system.

For an infinite system, fluctuations and correlations are neglected, and become important as the system approaches the critical point, where the correlation length diverges and there is power-law divergence of some quantities (for example, for the Smoluchowski model

$$M_2(t) = \frac{M_2(t_0)}{1 - CM_2(t_0)t} ).$$

With decreasing size of the system, fluctuations and correlations become more important (Gruyer et al, 2013). There is no divergence of the second moment  $M_2(t)$  (because such singularities are only possible within infinite systems), but it is expected to reach a maximum for a time close  $T_{gel} = [CM_2(t_0)]^{-1}$ . For this kind of systems there is an entire region with large fluctuations (in the vicinity of  $T_{gel} = [CM_2(t_0)]^{-1}$ ): the “pseudo critical region”.

**References:**

Bhattacharjee, S. M. (2001). Critical Phenomena: An Introduction from a modern perspective. In *Field Theories in Condensed Matter Physics* (pp. 69-117). Hindustan Book Agency, Gurgaon.

**Reviewer:**

15. L154: What is “product kernel”? If it is widely used, please provide several references.



*What are the assumptions for the kernel, linear drag, gravity only? Could you please explain why you choose this kernel?*

**Reply:**

The product kernel is a kernel proportional to the product of the masses of the colliding particles  $K(i, j) = Cx_i x_j$ . It is widely used because analytical solutions of the kinetic collection (KCE) or Smoluchowski equation (Eq. 1) have been obtained for this kernel Golovin (1963), Scott (1968), Drake (1972) and Drake and Wright (1972). Additionally, analytical solutions have also been obtained for the constant and sum (with probability of collision proportional to the sum of the masses of the colliding particles) kernels.

Lushnikov (1978, 2004) and Tanaka and Nakazawa (1993) also obtained analytical solutions of the master equation for the product kernel. Also, it is widely known that the product kernel is a gelling kernel (Lushnikov, 1978; 2004), and Lushnikov (1978, 2004) analytically obtained post gel particle size distributions for this case.

The aforementioned factors explain why we chose this kernel: It is a gelling kernel, and (due to the existence of analytical solutions), there is analytical expression for the sol gel transition time ( $T_{gel} = [CM_2(t_0)]^{-1}$ ) for this case. Then, it served as a benchmark for our Monte Carlo experiments, and to evaluate our method for the calculation of the pseudo-critical region.

The value of the constant  $C$  ( $C=5.49 \times 10^{10}$ ) in the product kernel  $K(i, j) = Cx_i x_j$  is a result of the polynomial approximation (Long, 1974):

$$K(x, y) = A + B(x + y) + Cxy \quad (S9)$$

of the hydrodynamic collection kernel:

$$K(x, y) = \pi [R(x) + r(y)]^2 E(x, y) [V(x) - V(y)] \quad (S10)$$

Long (1974) calculated the coefficients for the polynomials (9) approximating the hydrodynamic kernel (10) when the largest of the colliding drops is smaller than 50  $\mu\text{m}$ . The results obtained by Long (1974) are displayed in the table S1 (Alfonso et al, 2008).

**Table S1. Polynomials approximating the actual collection kernel  $K(x,y)$  (Long, 1974).**

Approximating Polynomial $P(x,y)$	Coefficients $R \leq 50 \mu\text{m}$ ( $\text{cm}^3 \text{sec}^{-1}$ )
$K(x, y) = A$	$A=1.20 \times 10^{-4}$
$K(x, y) = A + B(x + y)$	$A=0$ $B=8.83 \times 10^2$
$K(x, y) = Cxy$	$C=5.49 \times 10^{10}$
$K(x, y) = A + B(x + y) + Cxy$ $A=B^2/C$	$A=4.41 \times 10^{-7}$ $B=1.36 \times 10^2$

	$C=4.18 \times 10^{10}$
$K(x, y) = A + B(x + y) + Cxy$	$A=0$ $B=4.16 \times 10^2$ $C=2.24 \times 10^{10}$

### References:

- Drake, R.L.: The scalar transport equation of coalescence theory: Moments and kernels, *J. Atmos. Sci.*, 29, 537-547, 1972.
- Drake, R.L., Wright, T.J.: The scalar transport equation of coalescence theory: New families of exact solutions, *J. Atmos. Sci.*, 29, 548-556, 1972.
- Golovin, A.M.: The solution of the coagulating equation for cloud droplets is a rising air current, *Bull. Acad. Sci. USSR, Geophys. Ser.*, No. 5, 482-487, 1963.
- Long, A. B. (1974). Solutions to the droplet collection equation for polynomial kernels. *Journal of the Atmospheric Sciences*, 31(4), 1040-1052.
- Lushnikov, A. A.: Coagulation in finite systems. *Journal of Colloid and Interface Science*, vol. 65, no 2, p. 276- 285, 1978.
- Lushnikov, A. A.: From sol to gel exactly. *Physical review letters*, vol. 93, no 19, p. 198302, 2004.
- Scott, W.T.: Analytic studies of cloud droplet coalescence, *J. Atmos. Sci.*, 25, 54-65, 1968.
- Tanaka, H., Nakazawa, K.: Stochastic coagulation equation and the validity of the statistical coagulation equation, *J. Geomag. Geoelectr.*, 45, 361-381, 1993.

### Reviewer:

16. L164: Could you please explain what kinds of “Monte Carlo algorithm” you used? Is it comparable to Shima et al. 2009 [15], Li et al (2017) [12] and Li et al (2018) [13]? I understand you focus on the collision-coalescence process of cloud droplets. Could you please also provide the equations you solved numerically? Also, can you explain the difference of your “Monte Carlo algorithm” with those of Shima et al. 2009 [15], Li et al (2017) [12] and Li et al (2018) [13].

### Reply:

In this study we use the stochastic algorithm developed by Gillespie (1976) for chemical reactions. This algorithm was reformulated to simulate the kinetic behaviour of aggregating systems by Laurenzi and Diamond (1999), by defining species as a type of aggregate with a specific size and composition. In our case, species represent droplets of different sizes.

**As was remarked in the reply to question (2)**, the main difference between the Gillespie’s stochastic simulation algorithm (SSA) and other Monte Carlo methods based on the simulation particles (SIP) approach (like the Super Droplet method developed by Shima et al. (2009)), is that the Gillespie’s SSA involved the collision of only two physical particles (droplets in our case) per MC cycle, while in the approach based on SIP in each MC cycle collide SIP (super-droplets, for example) that represents a multiple number of droplets with the same attributes (radius  $r$  or mass in the simplest case) and position.

Could you please also provide the equations you solved numerically?

The **Gillespie (1976) algorithm** generates a statistically correct trajectory (possible solution) of the master equation. The steps below summarize the algorithm:

- 1) **Initialization:** Initialize the number of droplets in each species (the species are defined as droplets of different sizes). There is a unique index  $\mu$  for each pair of droplets  $i, j$  that may collide. For a system with  $N$  species  $(n_1, n_2, \dots, n_N)$

$\mu \in \frac{N(N+1)}{2}$ . The set  $\{\mu\}$  defines the total collision space, and is equal to the total number of possible interactions.

- 2) **Monte Carlo step:** Generate random numbers to determine the next collision to occur, as well as the time to the next collision. The next collision  $\mu$  is calculated according to

the distribution  $P_2(\mu) = \frac{a_\mu}{\alpha}$ , where  $a_\mu$  are calculated from the probabilities:

$a(i, j) = V^{-1}K(i, j)n_i n_j dt \equiv \text{Pr}\{\text{Probability that two unlike particles } i \text{ and } j \text{ with populations (number of particles) } n_i \text{ and } n_j \text{ will collide within the imminent time interval}\}$

$a(i, i) = V^{-1}K(i, i)\frac{n_i(n_i - 1)}{2} dt \equiv \text{Pr}\{\text{Probability that two particles of the same species } i \text{ with population (number of particles) } n_i \text{ collide within the imminent time interval}\}$

and  $\alpha = \sum_{\nu=1}^{\frac{N(N+1)}{2}} a_\nu$ . The time to the next collision is exponentially distributed with mean  $1/\alpha$

- 3) **Update:** Increase the time by the randomly generated time in Step 2. Update the droplet count based on the collision that occurred.  
 4) **Iterate:** Go back to Step 2 unless the number of droplets is zero or the simulation time has been exceeded.

In the revised version, the details of the Monte Carlo method will be discussed in one appendix.

**Reviewer:**

17. L167: Can you give a physical explanation about why you choose “ $C=5,49 \cdot 10^{10} \text{ cm}^3 \text{ s}^{-1}$ ”?

**Reply:**

(This was discussed in the answer to question 15):

The value of the constant  $C$  ( $C=5.49 \times 10^{10}$ ) in the product kernel  $K(i, j) = Cx_i x_j$  is a result of the polynomial approximation (Long, 1974):

$$K(x, y) = A + B(x + y) + Cxy \quad (\text{S9})$$

of the hydrodynamic collection kernel:

$$K(x, y) = \pi [R(x) + r(y)]^2 E(x, y) [V(x) - V(y)] \quad (\text{S10})$$

Long (1974) calculated the coefficients for the polynomials (9) approximating the hydrodynamic kernel (10) when the largest of the colliding drops is smaller than 50  $\mu\text{m}$ . The results obtained by Long (1974) for the product kernel are displayed in Table S1 (see the answer to question 15).

**Reviewer:**

18. L173: Could you please explain more about the “mixing fraction”, like mixing fraction of which quantity and the corresponding physical picture or intuition?

**Reply:**

Looking for more clarity, we will change the terminology. In the equation (6) of the paper:

$$f(x, \theta, \mu_1, \beta, \mu_2, \sigma) = \theta \text{Gumbel}(x, \mu_1, \beta) + (1 - \theta) \text{Gauss}(x, \mu_2, \sigma)$$

the coefficients  $\theta$  and  $(1 - \theta)$  are the probabilities associated with each component and are called the **mixture weights**. The **individual distributions**  $\text{Gumbel}(x, \mu_1, \beta)$  and  $\text{Gauss}(x, \mu_2, \sigma)$  that are combined to form the mixture, are the mixture components. We **will eliminate the term “mixing fraction”** in the revised version looking for more clarity.

**Reviewer:**

20. L260: Please provide reference for “The block maxima (BM) approach in extreme value theory (EVT) was applied” and compare with the large deviation theory/method described in Wilkinson 16 [3].

**Reply:**

The main (classical) reference for extreme value theory (EVT) and the block maxima approach in is the book of **Gumbel (1958)**.

The large deviation theory (LDP) is concerned with the behavior of the tails of the distribution for the sum of independent and identical distributed (i.i.d) random variables. The Central Limit theorem is limited to values of the random variable not too far from the mean value. In the revised version, the difference between the Wilkinson’s (2016) approach and ours will be discussed in both the introduction and conclusions. The discussion below will add more clarity to the discussion of the differences between the two approaches.

**Large deviation theory (LDT):** Suppose  $X_1, X_2, \dots, X_n$  is a sequence of independent, identically distributed random variables, with mean  $\mu$  and variance  $\sigma^2$ , and  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ , then:

$$\lim_{n \rightarrow \infty} \Pr\left(\frac{\bar{X}_n - \mu}{\sigma\sqrt{n}} \leq z\right) = \Phi(z) \quad (\text{S10})$$

Then, for large  $n$ , the random variable  $\frac{\bar{X}_n - \mu}{\sigma\sqrt{n}}$  have a Gaussian distribution with mean  $\mu=0$  and standard deviation 1. This is the Central Limit theorem (CLT). Then, the CLT says that

$$\lim_{n \rightarrow \infty} \Pr\left(\bar{X}_n \leq \mu + z\sigma\sqrt{n}\right) = \Phi(z) \quad (\text{S11})$$

And

$$\lim_{n \rightarrow \infty} \Pr\left(\bar{X}_n \geq \mu + z\sigma\sqrt{n}\right) = 1 - \Phi(z) \quad (\text{S12})$$

However, we must take into account that the range of validity of the CLT is:

$$|x - \mu| = O\left(\frac{1}{\sqrt{n}}\right) \quad (\text{S12})$$

And as a consequence, is not very accurate in the tails of the distribution. For example, if we approximate the tail probability  $\Pr(\bar{X}_n \geq \mu + \varepsilon)$  for a fixed value of  $n$  by using (S12), the result will be not very accurate if  $\bar{X}_n$  is far from the mean, unless  $n$  is sufficiently large. Then, we need an expression more accurate than the Gaussian distribution for finite but large values of  $n$ , and that will recover the Gaussian distribution when  $n \rightarrow \infty$ . The large deviation theory provides a solution for this problem. **Summarizing, the large deviation theory (LDP) is concerned with the behavior of the tails of the distribution. The Central Limit theorem is limited to values of the random variable not too far from the mean value. Then, according to the large deviation theory for sums of independent random variables:**

$$f(x) \rightarrow C(n)e^{-nI(x)} \quad (\text{S13})$$

Where  $C(n)$  is a normalizing constant, and  $I(x)$  is the large deviation function. The distribution (S13). This expression is accurate for all values of  $x$ , while the Gaussian is accurate only for values that fulfilled the condition  $|x - \mu| = O\left(\frac{1}{\sqrt{n}}\right)$ .

**Wilkinson (2016) application of large deviation theory (LDT):** Due to the fact that the precipitation occurs on a time scale that is smaller than the typical scale for one collision, the problem can be solve by applying large deviation theory (Wilkinson, 2016). Then, it is

necessary to determine the probability density for the time  $T_{N_c} = \sum_{i=1}^{N_c} t_i$  being a small fraction of its average value (far from the mean value). In the former expression the  $t_i$  are the times between droplet collisions and  $N_c$  the number of collisions, which have an exponential distribution (Wilkinson, 2016). Then, we need to find the probability at the tails of the distribution. According to LDT, the probability can be written in the form:

$$f(\tau) = \frac{1}{\langle T \rangle} e^{-J(\tau)} \quad (\text{S14})$$

where  $J(\tau)$  is the rate function (see Eq. S13) that was explicitly calculated in Wilkinson's (2016) paper.

**This paper's approach:** The approach we follow in this report is different, since we are interested on finding the size distribution of runaway droplets that trigger precipitation formation.

**References:**

- Gumbel, E.J.: Statistics of Extremes. Columbia University Press, 1958 - 375 pp.  
 Wilkinson M. 2016. Large deviation analysis of rapid onset of rain showers. Phys. Rev. Lett. 116: 018 501, doi:10.1103/PhysRevLett.116.018501.

**Reviewer:**

21. L271-272: *Please rephrase the sentence "The sample size...of data" to improve the readability. The "which clause" is not encouraged in scientific writing.*

**Reply:**

In the revised version the sentence will be rewritten as suggested.

**Reviewer:**

22. L318: *Did you mean "entire dataset"?*

**Reply:**

Yes, the "entire dataset". The corresponding change will be made.