Mexico City, August 3, 2015

This document includes:

i) A point-by-point response to the reviewers.

ii) A list of all relevant changes made in the manuscript

iii) A marked-up manuscript version showing the changes made in the manuscript

i) A point-by-point response to the reviewers.

First, I would like to thank the anonymous referees for their comments that will improve the quality of the paper. Our revised version will include several of their suggestions.

Anonymous referee # 1: **"No changes are advised. The overall strategy of using problem-specific constraints to overcome the curse of dimensionality is reminiscent of the method purportedly used to break the Enigma code according to the movie The Imitation Game. Thus it appears to be common knowledge, so there should be a citable reference. I don't know of any, but if the author can find one, it might be cited."**

Response to Anonymous Referee # 1

The curse of dimensionality (Bellman, R.; 1961) is a problem of both the data and the algorithm being applied. Solutions to this problem involve changing the algorithm or processing the data into a lower dimensional form. This can be done in many problems since high dimensional data sets can be reduced to lower-dimensional without significant information loss. For example, in Bayesian statistics where posterior distributions have multiple dimensions, this problem was overcome with the implementation of the Markov Chain Monte methods. A new reference is added.

Reference:

Bellman, R. (1961), Adaptive Control Processes: A Guided Tour, Princeton University Press.

Anonymous Referee # 2.

Anonymous referee # 2: **"The author claims that the major advantage of his method is that it is the first one to account for correlations in probabilities of system states (paragraphs near the end of Sec. 1). However, a numerical method developed by Gillespie (Gillespie D. T., J.Atmos. Sci. 32 (10), 1977-1989, 1975), contrary to what the author suggests, also accurately reproduces master equation. In Gillespie's method, the master equation is not solved directly, but a single stochastic trajectory following this equation is obtained. Temporal evolution of probability distributions of different states can be obtained by averaging over many runs. Seeßelberg proposed a variation of Gillespie's method in which droplet mass is discretized into bins, which makes the algorithm applicable to significantly larger systems than the one proposed in the discussed paper (Seeßelberg et al., Atmospheric Research 40(1), 33-48, 1996). In my opinion a comparison of results and efficiency with Gillespie's method is necessary.**

The author should clearly outline what are the advantages of his method over the Gillespie's one and in which situations should it be used."

Response to Anonymous Referee # 2

All the referee's observations regarding the Stochastic Simulation Algorithm (SSA) of Gillespie (1975a) **will be incorporated in the revised version of the paper.**

In our paper we did just want to remark the fact that in Gillespie (1972), the evolution equation for the probability of finding a given number of *m*-drops of a particular size at time *t*, does not allow to solve for $P(n,m;t)$. This is due to the fact that the right hand side contains a set of unknown conditional probabilities (Gillespie, 1972). Then, the simplest way to proceed in order to close the system was to ignore the correlations.

1. **Comparison between the Gillespie's SSA and the numerical algorithm.**

As we know, in the Gillespie's SSA, the ensemble mean for the number of droplets of each droplet mass is calculated from the expression (Gillespie, 1975a):

$$N(m;t) = \frac{1}{N_r}\sum_{i=1}^{N_r} N^i(m;t) \tag{1}$$

where N_r is the number of realizations of the stochastic algorithm, $N^i(m;t)$ is the number of droplets of mass *m* in the *i*-realization at time *t*, and $N(m;t)$ is the ensemble mean. From expression (1), it is clear that in order to obtain the correct expected values $(N(m;t))$ at the large end of the droplet size distribution; we will need a huge number of realizations of the stochastic algorithm (> 10^9).

To further investigate this question, the evolution of a cloud system with an initial monodisperse droplet size distribution of N_0=30 droplets of 14 *μm* in radius (droplet mass 1.1494×10^{-8}g) at t_0, and a volume of 1cm^3 was calculated with both the numerical algorithm and the SSA of Gillespie.

The results obtained by the two methods, were then compared with the analytical solution of the master equation (Eq.(13) our paper) obtained by Tanaka and Nakazawa (1993) for the same conditions.

The averages calculated from the Gillespie's method for N_r=10^3 realizations, and the analytical solution at *t*=1200 are displayed in Fig. 1. As can be observed both the Monte Carlo averages and the analytical solution are closely coincident for the small end of the droplet size distribution. However, due to the small number of realizations, the SSA fails to reproduce the distribution for the expected values at the large end (See Table 1).

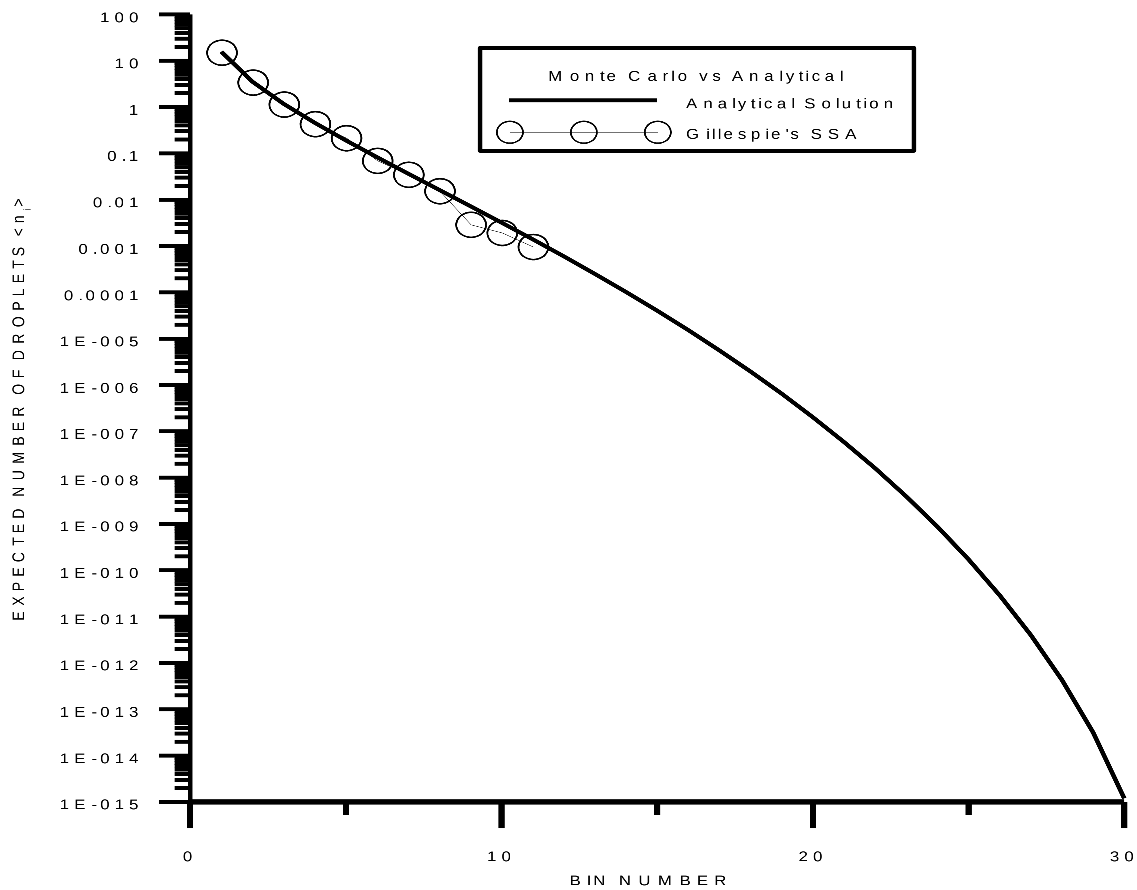


FIG. 1. For the sum kernel, size distribution obtained from the analytical solution of the master equation (line) and the SSA of Gillespie for 10^3 realizations (circles) at *t*=1200 sec. Calculations were performed with the initial condition $P(30,0,0,0,...,0;0)=1$ **and the sum kernel** $K(i,j)=B(x_i+x_j)$ **, with *B*=8.82×10^2 cm^3 sec^{-1}.**

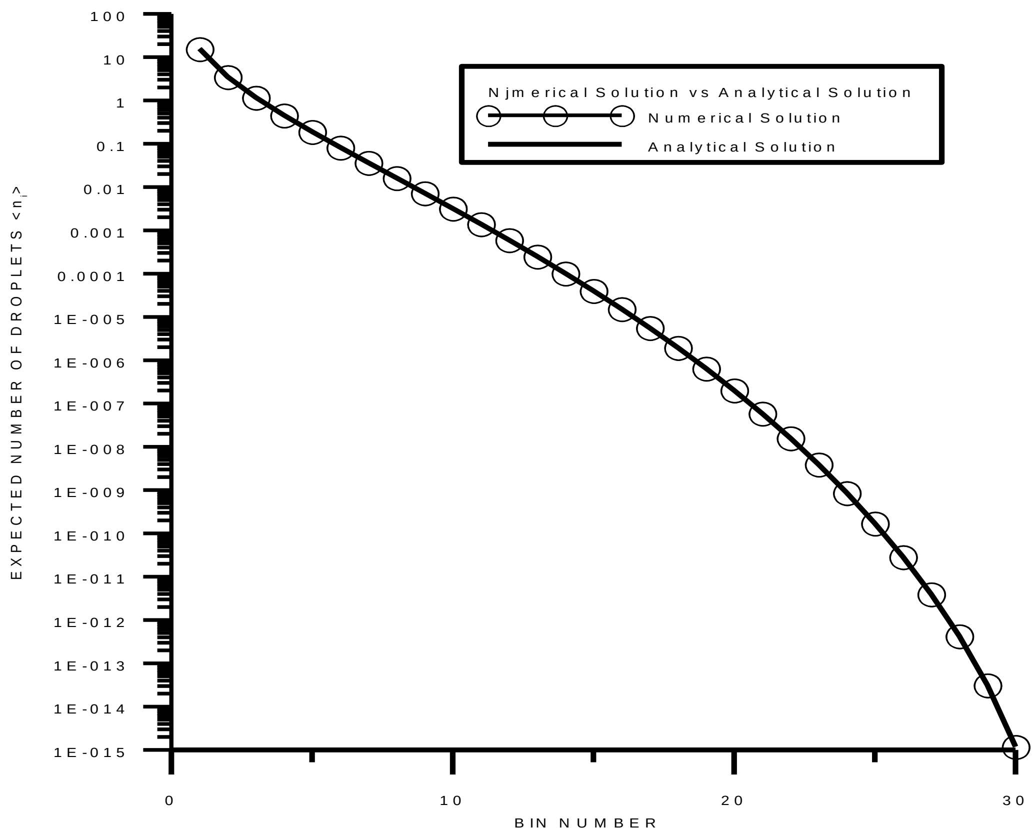


FIG. 2. For the sum kernel, size distribution obtained from the analytical solution of the master equation (line) and the numerical algorithm proposed in this paper (circles) at *t*=1200 sec. Calculations were performed with the initial condition $P(30,0,0,0,...,0;0)=1$ **and the sum kernel** $K(i,\, j)=B(x_i+x_j)$ **, with *B*=8.82×10^2 cm^3 sec^{-1}.**

For a more detailed analysis, the expected number of particles for each droplet size calculated from the analytical solution, the numerical algorithm and the SSA of Gillespie (for 1000 and 10,000 realizations) are displayed in Table 1. As can be checked in the table, the size distributions are almost identical for the small end, but they differ substantially at the large end since the SSA produces no particles larger than $12v_0$ and $16v_0$ for 1000 and 10,000 realizations respectively (v_0 =1.1494×10^{-8}g, mass of a 14 *μm* droplet).

For 1000 realizations, the Monte-Carlo averages differ from the analytical solution for bin numbers larger than 8. For 10,000 realizations we have the same situation for bin numbers larger than 13.

Table 1. Expected values for each droplet mass obtained at *t*=1200 sec. for the analytical solution, the numerical algorithm proposed in this work, and the Gillespie's SSA (for 1000 and 10,000 realizations). Calculations were performed with the initial condition $P(30,0,0,0,...,0;0)=1$ and the sum kernel $K(i,j)=B(x_i+x_j)$, with *B*=8.82×10^2 cm^3 sec^{-1}.

Expected values for each droplet size: $\langle n_i \rangle$

t=1200 sec.

Bin Number	Analytical Solution	Numerical algorithm	SSA (N_r= 1000)	SSA (N_r= 10,000)
1.000	1.5633E+01	1.5622E+01	1.5612E+01	1.5619E+01
2.000	3.5302E+00	3.5303E+00	3.5250E+00	3.5425E+00
3.000	1.1754E+00	1.1762E+00	1.1870E+00	1.1712E+00
4.000	4.5543E-01	4.5609E-01	4.4800E-01	4.5050E-01
5.000	1.9017E-01	1.9057E-01	2.2300E-01	1.9600E-01
6.000	8.2592E-02	8.2824E-02	7.2000E-02	8.2000E-02
7.000	3.6583E-02	3.6709E-02	3.6000E-02	3.6800E-02
8.000	1.6320E-02	1.6387E-02	1.6000E-02	1.6100E-02
9.000	7.2696E-03	7.3034E-03	3.0000E-03	6.5000E-03
10.000	3.2117E-03	3.2284E-03	2.0000E-03	3.5000E-03
11.000	1.3997E-03	1.4077E-03	1.0000E-03	1.2000E-03
12.000	5.9891E-04	6.0263E-04	0.0000E+00	4.0000E-04
13.000	2.5049E-04	2.5216E-04	0.0000E+00	4.0000E-04
14.000	1.0197E-04	1.0269E-04	0.0000E+00	3.0000E-04
15.000	4.0229E-05	4.0529E-05	0.0000E+00	0.0000E+00
16.000	1.5312E-05	1.5431E-05	0.0000E+00	1.0000E-04
17.000	5.5954E-06	5.6404E-06	0.0000E+00	0.0000E+00
18.000	1.9526E-06	1.9687E-06	0.0000E+00	0.0000E+00
19.000	6.4672E-07	6.5217E-07	0.0000E+00	0.0000E+00
20.000	2.0189E-07	2.0361E-07	0.0000E+00	0.0000E+00
21.000	5.8917E-08	5.9419E-08	0.0000E+00	0.0000E+00
22.000	1.5913E-08	1.6048E-08	0.0000E+00	0.0000E+00
23.000	3.9295E-09	3.9622E-09	0.0000E+00	0.0000E+00
24.000	8.7349E-10	8.6634E-10	0.0000E+00	0.0000E+00
25.000	1.7127E-10	1.7176E-10	0.0000E+00	0.0000E+00
26.000	2.8809E-11	2.8765E-11	0.0000E+00	0.0000E+00
27.000	3.9922E-12	3.9906E-12	0.0000E+00	0.0000E+00
28.000	4.2746E-13	4.2803E-13	0.0000E+00	0.0000E+00
29.000	3.1450E-14	3.1525E-14	0.0000E+00	0.0000E+00
30.000	1.1930E-15	1.1962E-15	0.0000E+00	0.0000E+00

As expected, for 1000 and 10000 realizations, no states with droplets 30 times larger than monomer-sized ones were realized. A minimum number of N_r

realizations must be performed in order to obtain expected values larger or equal to 10^{-Nr}.
At the same time, the numerical algorithm performed very well, with expected values that are very close to the analytical solution.
We can conclude that our method will be suitable if we need to accurately calculate the large end of the droplet spectrum for small systems, with <50 monomer droplets in the initial state. As for that case the SSA requires a large number of realizations, it will be computationally very expensive. Then, our method will be a good alternative, as it provides the desired accuracy to detect the possible small differences between different numerical approaches. It can also work as a benchmark for different Monte Carlo methods for the collision coalescence process.
A comparison of results and efficiency with Gillespie's method will be included in the revised version of the paper.
Anonymous referee # 2: **"Other thing that would be beneficial is a short discussion of how do results presented in the paper actually relate to cloud development. From Figs. 6, 7 and 9 one can conclude that fluctuations and correlations tend to delay formation of large droplets in small volumes (in comparison with the deterministic equation). However, these results cannot be extrapolated to larger systems. Such a small volume would not remain undisturbed for time of the order of 1000s. Does the author expect that fluctuations in small volumes can have significant impact on development of clouds."**

2. On Seeβelberg's variation of Gillespie's SSA.

In its original version, the Gillespie's Stochastic Simulation Algorithm (Gillespie, 1975a) was formulated as a particle accounting algorithm, as it keeps track of every individual particle in the system. Unfortunately, this type of algorithm can be computationally very demanding, as a system of N particles requires storage of $N(N+1)/2$ transition probabilities. Then, although it yields exact realizations of the coalescence process, the requirements of computational storage strongly limit its applicability.
The algorithm of Seeβelberg's et al. (1996) found a way to overcome this problem by dividing the drop population into classes and storing only the transitional probabilities between classes. The problem with that approach was the method of removal of the coalescing drops from "colliding classes", a process that introduces an error.
The numerical difficulties of Gillespie (1975a) were significantly overcome in the modified version proposed by Laurenzi et al. (2002). Within their approach, they define collisions between species (that are hydrometeors with the same attributes: mass and composition).By using this framework, there is only the need to store the probabilities of "aggregation reaction" between species, with a considerable reduction of computer memory requirements. After a collision, the reactant (colliding) and product species are updated without any approximation, and not any kind of systematic errors are introduced.

3. On the applicability of the finite volume approach

In defense of the finite system approach, it might be argued that, in the early stages of precipitation formation, due to small terminal velocities of the droplets, the coalescence process is a fairly localized process. Then, two droplets in widely separated parts of the cloud are not going to be coalescing with each other. This was the approach followed by Bayewitz et al. (1974) (and endorsed in Gillespie, 1975a). In their paper, for comparing the stochastic and kinetic approaches, they partitioned the cloud into many sub-volumes, with no collisions being permitted for two droplets of different sub-volumes.

A more complex model that uses the master equation formalism, and introduces the interactions between the sub-volumes was developed by Merkulovich and Stepanov (1990, 1994). This model is based on a scheme proposed by Nicolis and Prigogine (1984). Within this theory, the whole system is subdivided into subsystems that can be considered spatially homogeneous, and interactions between neighbors occur through particle exchange. That leads to a very complex set of master equations for each sub-volume. Although very complex, it could be a starting point in order to consider the interactions between cells through sedimentation or other processes.

4. **On the impact of fluctuations on cloud development**

As was discussed, in systems of small populations statistical fluctuations become important and the outcome from kinetic equations may differ from the stochastic means. However, fluctuations will also be very important when the system is near a critical point. In the cloud physics context, a critical point could be related with the formation of a droplet with mass comparable to the mass of the initial system (Alfonso et al., 2013; Lushnikov, 2004). This could be a mechanism responsible for the formation of droplet embryos that trigger precipitation formation.

At this moment, fluctuations will have a significant impact in the development of the system and the total mass predicted by the KCE will start to decrease. This is usually interpreted to mean that a super particle has formed (known as a *gel*) and the system exhibits a sol-gel transition (also called *gelation*).

Because the master equation employs the stochastic approach without any approximations, it can predict the behavior of the collection process at all times. This way, the expected values at the large end of the droplet size distribution can be calculated after the total mass predicted by the KCE starts to decrease (Lushnikov, 2004). By using this method, an accurate comparison between the kinetic and stochastic methods can be performed after the super particle is formed. Then, for this case, it is expected to obtain broader droplet mass distributions by using the stochastic approach.

All this questions will be discussed in the revised version of the paper, as suggested by the referee. A follow-up paper (that is under preparation) will be devoted to a more detailed analysis of all this problems.

5. **Minor comments:**

a) **Referee Comment:** In Section 2, l. 14 author writes that to solve master equation directly, arrays of the size of 3 x 10^{20} elements would have to be used, where does this number come from?

Answer to the comment:
The number of elements in the array will be 1.34 x 10^{16}, and not 3 x 10^{20} as was written in the paper. **This number will be corrected in the revised version**. The explanation follows:
If we attempt to solve the master equation by brute force by using a typical discretization, then, for an initial monodisperse droplet distribution with *N* monomers, we will need to define an *N*-dimensional array (due to the fact that we have configurations of the form $P(n_1, n_{2,}, ..., n_k, ...; t)$). If we have *n1* =N monomers of size k=1 at t=0, then, the maximum number of particles for size k=2 will be [*N*/2], for size k=3, [*N*/3], and so on, where [*N*/k] denotes the closest smaller integer. (For example, the maximum possible number of droplets of size k=*N* is [*N*/*N*]=1). Of course, due to the mass conservation relation:

$$\sum_{i=1}^{N} x_i n_i = M_T \tag{2}$$

these maximum values for each droplet size can be attained only if the number of droplets for the rest of the bins are set equal to zero or close to zero. The number of elements of the resulting array then will be the product: $N*[N/2]*[N/3]*[N/(N-1)]*[N/N]$. If evaluate this expression for *N*=50, the value of 1.34x10^{16} will be obtained.

b) **Referee Comment:** - In the Abstract it should be more clearly stated that the kinetic collection equation and Smoluchowski coagulation equation are different names for the same equation.

Answer to the comment: The corresponding correction will be made in the revised version as suggested.

6. Technical Comments:

The corresponding modifications will be made.

7. References

Alfonso, L., Raga, G.B., Baumgardner, D.: The validity of the kinetic collection equation revisited. Part II: Simulations for the hydrodynamic kernel, *Atmos. Chem. Phys.*, 10, 6219-6240, 2010.

Alfonso, L., Raga, G. B., 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.

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.

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

Gillespie, D.T.: An Exact Method for Numerically Simulating the Stochastic Coalescence Process in a Cloud, *J. Atmos. Sci*. 32, 1977-1989, 1975a.

Laurenzi, I. J., Bartels, J. D., & Diamond, S. L. (2002). A general algorithm for exact simulation of multicomponent aggregation processes. *Journal of Computational Physics*, *177*(2), 418-449.

Lushnikov, A. A.: From sol to gel exactly. *Physical review letters*, vol. 93, no 19, p. 198302, 2004.

Merkulovich, V. M., and A. S. Stepanov. "Statistical description of coagulation in finite spatially inhomogeneous systems." *Atmospheric research* 25.5 (1990): 431-444.

Merkulovich, V. M., and A. S. Stepanov. "Statistical description of coagulation in finite spatially inhomogeneous systems (Part 2)." *Atmospheric Research* 26.4 (1991): 311-327.

Nicolis, Grégoire, and M. Malek Mansour. "Onset of spatial correlations in nonequilibrium systems: a master-equation description." *Physical Review A* 29.5 (1984): 2845.

Valioulis, I.A., List, E.J.: A numerical evaluation of the stochastic completeness of the kinetic coagulation equation, J. Atmos. Sci., 41, 2516-2529, 1984.

ii) A list of all relevant changes made in the manuscript

Modifications to paper ACP-2015-144

1. **Reviewer # 2:** "In the Abstract it should be more clearly stated that the kinetic collection equation and Smoluchowski coagulation equation are different names for the same equation."

Now this fact is clearly stated in the abstract:

"In cloud modeling studies, the time evolution of droplet size distributions due to collision-coalescence events is usually modeled with the Smoluchowski coagulation equation, also known as the kinetic collection equation (KCE)".

2. **Reviewer # 2: "**The author claims that the major advantage of his method is that it is the first one to account for correlations in probabilities of system states (paragraphs near the end of Sec. 1). However, a numerical method developed by Gillespie (Gillespie D. T., J. Atmos. Sci. 32 (10), 1977-1989, 1975), contrary to what the author suggests, also accurately reproduces master equation".

The corresponding additions were made in section 1 following referee's suggestion:

"It is noteworthy to mention that the Stochastic Simulation Algorithm (SSA) developed by Gillespie (1975) also accurately reproduces the master equation. In Gillespie's method, the master equation is not solved directly, but a statistically correct trajectory (possible solution) of the master equation is generated. At any time, expected values at each droplet size can be obtained by averaging over many runs. However, a large number of realizations are necessary in order to obtain the desired

accuracy at the large end of the droplet size distribution. A detailed comparison between the two methods will be made in section 3".

3. **Reviewer # 2: "**In my opinion a comparison of results and efficiency with Gillespie's method is necessary. The author should clearly outline what are the advantages of his method over the Gillespie's one and in which situations should it be used."

   In the new, revised version, following referee's suggestion a new section is added **at page 10 of the manuscript** (*3.2 Comparison with the SSA of Gillespie.*) with a detailed comparison of the numerical method with Gillespie's stochastic simulation algorithm:

3.2 Comparison with the SSA of Gillespie.

"As was mentioned in the introduction, the algorithm of Gillespie generates a statistically correct trajectory of the stochastic master equation. It was presented in Gillespie (1975), and popularized in Gillespie (1977) were it was used to simulate chemical systems. As we know, in Gillespie's SSA, the ensemble mean for the number of droplets at each droplet mass is calculated from the expression (Gillespie, 1975):

$$N(m;t) = \frac{1}{N_r}\sum_{i=1}^{N_r} N^i(m;t) \qquad \textbf{(14)}$$

where N_r is the number of realizations of the stochastic algorithm, $N^i(m;t)$ is the number of droplets of mass m in the i-realization at time t, and $N(m;t)$ is the ensemble mean. From expression (14) it is clear that in order to obtain the correct expected values $(N(m;t))$ at the large end of the droplet size distribution; we will need a large number of realizations of the SSA.

To further investigate this question, the evolution of a cloud system with an initial mono-disperse droplet size distribution of N_0=30 droplets of 14 µm in radius (droplet mass 1.1494×10^{-8}g) at t_0, and a volume of $1cm^3$ was calculated with both the numerical algorithm and the Gillespie's SSA for the sum kernel ($K(i,j) = B(x_i + x_j)$, with $B=8.82\times10^2$ cm^3 sec^{-1}) . The results obtained by the two methods, were then compared with the analytical solution of the master equation (Eq. (13)) obtained by Tanaka and Nakazawa (1993) for the same conditions.

The averages calculated from the Gillespie's method for N_r=10^3 realizations, and the analytical solution at t=1200 are displayed in Fig. 5. As can be observed, both the Monte Carlo averages and the analytical solution are closely coincident for the small end of the droplet size distribution. However, due to the small number of realizations, the SSA fails to reproduce the distribution for the expected values at the large end (See Table 1).

For a more detailed analysis, the expected number of particles for each droplet size calculated from the analytical solution, the numerical algorithm and the SSA of Gillespie (for 1000 and 10,000 realizations) are displayed in Table 1. As can be checked in the table, the size distributions are almost identical for the small end. However, they differ substantially at the large end since the SSA produces no particles larger than 12v_0 and 16v_0 for 1000 and 10,000 realizations respectively (v_0 =1.1494×10^{-8}g, mass of a 14 µm droplet).

For 1000 realizations, the Monte-Carlo averages differ from the analytical solution for bin numbers larger than 8. For 10,000 realizations we have the same situation for bin numbers larger than 13.

As expected, for 1000 and 10000 realizations, no states with droplets 30 times larger than monomer-sized ones were realized. The numerical algorithm described in this paper performed very well at the large end, with expected values that are very close to the analytical solution (see Fig5 and Table 3).

It can be concluded that our method will be suitable if we need to accurately calculate the large end of the droplet spectrum for small systems (with <50 monomer droplets in the initial state). As the SSA requires a large number of realizations, it will be computationally very expensive. Then, for a small number of particles, our algorithm will be a good alternative, as it provides the desired accuracy to detect the possible small differences between different numerical approaches. It can also work as a benchmark for different Monte Carlo methods for the collision coalescence process. "

4. **Reviewer # 2**: Other thing that would be beneficial is a short discussion of how do results presented in the paper actually relate to cloud development. From Figs. 6, 7 and 9 one can conclude that fluctuations and correlations tend to delay formation of large droplets in small volumes (in comparison with the deterministic equation). However, this results can not be extrapolated to larger systems. Such a small

volume would not remain undisturbed for time of the order of 1000s. Does the author expect that fluctuations in small volumes can have significant impact on development of clouds?

A short discussion on the applicability of the finite volume stochastic approach was added in section 5:

“Can be a topic of discussion the limits of applicability of the finite volume approach to problems of precipitation formation, since such small volumes would not remain undisturbed for a long time in a real cloud. However, in defense of the finite system approach, it might be argued that, in the early stages of cloud development, due to small terminal velocities of the droplets, the coalescence process is a fairly localized process. Then, two droplets in widely separated parts of the cloud are not going to be coalescing with each other. This was the approach followed by Bayewitz et al. (1974) (and endorsed in Gillespie (1975)). In their paper, for comparing the stochastic and kinetic approaches, they partitioned the cloud into many sub-volumes, with no collisions being permitted for two droplets of different sub-volumes. However, interactions between sub-volumes through sedimentation, diffusion or other physical processes were not considered.

For a constant collection kernel, a more complex model that uses the master equation formalism, and introduces the interactions between the sub-volumes was developed by Merkulovich and Stepanov (1990, 1991). This model is based on a scheme proposed by Nicolis and Prigogine (1977) for chemical reactions. Within this theory, the whole system is subdivided into sub-volumes (coalescence cells) that can be considered spatially homogeneous. Coalescence events are permitted only between droplets from the same sub-volume, and interactions between neighbors occur through the diffusion process. That leads to a set of master equations for each sub-volume. Although very complex, it could be a starting point in order to consider the interactions between small coalescence volumes through sedimentation or other physical mechanisms.

However, fluctuations will be also very important, if the collection kernel K(i,j) increases sufficiently rapidly with i and j and a giant droplet with mass comparable to the total mass of the system is formed. In that case, the total mass predicted by the KCE starts to decrease. This is usually interpreted to mean that the system exhibits a phase transition (also called gelation). After this moment, the true averages calculated

from the master equation will differ from the averages obtained from Eq. 1, and there is a transition from a system with a continuous droplet distribution to one with a continuous distribution plus a giant cluster (Alfonso et al., 2013). After the sol-gel transition the KCE breaks down: the second moment of the size distribution diverges at the gel point, and, as was remarked, the first moment decays, i.e., mass is not conserved.

The limitation of the KCE equation arises from the fact that it is a deterministic equation with no fluctuations or correlations included. Then it describes an inherently stochastic process with a single metric, the mean cluster distribution (Matsoukas, 2015). Then, in order to model properly the system behavior after the giant cluster is formed, the role of fluctuations should be considered.

By using the finite volume approach, the expected values at the large end of the droplet size distribution can be obtained in the post-gel region (Lushnikov, 2004; Matsoukas, 2015), and be compared with the expected values obtained from the kinetic approach. As a result, it is expected to obtain broader droplet mass distributions by using the stochastic approach. A follow-up paper will be devoted to a more detailed analysis of all these problems".

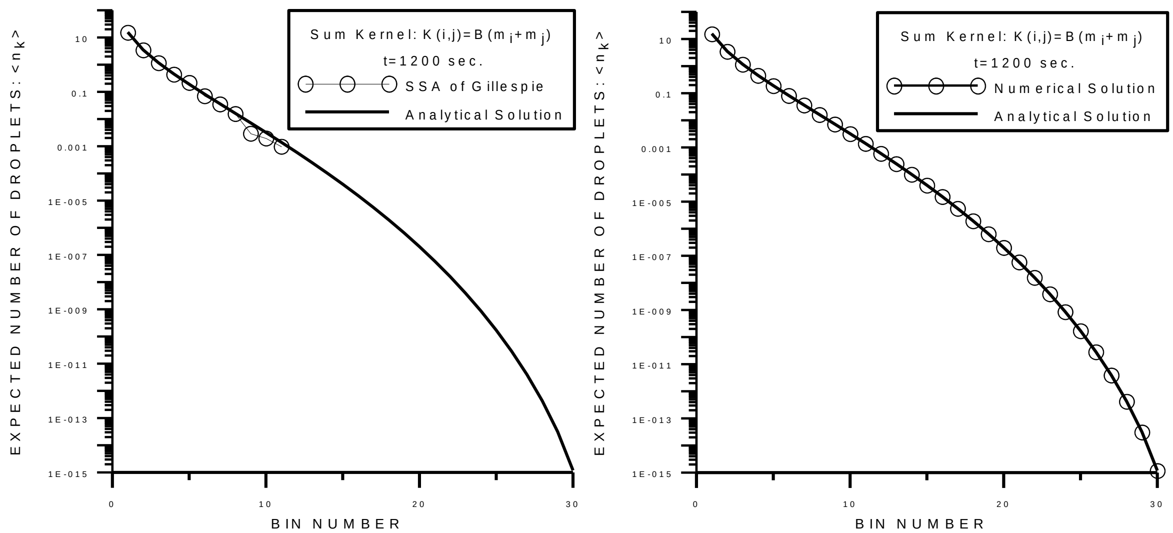
5. **New Table added to the paper:** Table 3 is now added to the revised version in section: *3.2 Comparison with the SSA of Gillespie.*

Table 3. Expected values for each droplet mass obtained at t=1200 sec. for the analytical solution, the numerical algorithm proposed in this work, and the Gillespie's SSA (for N_r= 1000, 10000 realizations). Calculations were performed with the initial condition $P(30,0,0,0,...,0;0)=1$, and the sum kernel $K(i,j)=B(x_i+x_j)$ with B=8.82×10^2 cm^3 sec^{-1}.

Expected values for each droplet size: $\langle n_i \rangle$, t=1200 sec.				
Bin Number	Analytical Solution	Numerical algorithm	SSA (N_r= 1000)	SSA (N_r= 10,000)
1.000	1.5633E+01	1.5622E+01	1.5612E+01	1.5619E+01
2.000	3.5302E+00	3.5303E+00	3.5250E+00	3.5425E+00

3.000	1.1754E+00	1.1762E+00	1.1870E+00	1.1712E+00
4.000	4.5543E-01	4.5609E-01	4.4800E-01	4.5050E-01
5.000	1.9017E-01	1.9057E-01	2.2300E-01	1.9600E-01
6.000	8.2592E-02	8.2824E-02	7.2000E-02	8.2000E-02
7.000	3.6583E-02	3.6709E-02	3.6000E-02	3.6800E-02
8.000	1.6320E-02	1.6387E-02	1.6000E-02	1.6100E-02
9.000	7.2696E-03	7.3034E-03	3.0000E-03	6.5000E-03
10.000	3.2117E-03	3.2284E-03	2.0000E-03	3.5000E-03
11.000	1.3997E-03	1.4077E-03	1.0000E-03	1.2000E-03
12.000	5.9891E-04	6.0263E-04	0.0000E+00	4.0000E-04
13.000	2.5049E-04	2.5216E-04	0.0000E+00	4.0000E-04
14.000	1.0197E-04	1.0269E-04	0.0000E+00	3.0000E-04
15.000	4.0229E-05	4.0529E-05	0.0000E+00	0.0000E+00
16.000	1.5312E-05	1.5431E-05	0.0000E+00	1.0000E-04
17.000	5.5954E-06	5.6404E-06	0.0000E+00	0.0000E+00
18.000	1.9526E-06	1.9687E-06	0.0000E+00	0.0000E+00
19.000	6.4672E-07	6.5217E-07	0.0000E+00	0.0000E+00
20.000	2.0189E-07	2.0361E-07	0.0000E+00	0.0000E+00
21.000	5.8917E-08	5.9419E-08	0.0000E+00	0.0000E+00
22.000	1.5913E-08	1.6048E-08	0.0000E+00	0.0000E+00
23.000	3.9295E-09	3.9622E-09	0.0000E+00	0.0000E+00
24.000	8.7349E-10	8.6634E-10	0.0000E+00	0.0000E+00
25.000	1.7127E-10	1.7176E-10	0.0000E+00	0.0000E+00
26.000	2.8809E-11	2.8765E-11	0.0000E+00	0.0000E+00
27.000	3.9922E-12	3.9906E-12	0.0000E+00	0.0000E+00
28.000	4.2746E-13	4.2803E-13	0.0000E+00	0.0000E+00
29.000	3.1450E-14	3.1525E-14	0.0000E+00	0.0000E+00
30.000	1.1930E-15	1.1962E-15	0.0000E+00	0.0000E+00

6. **New Figures added to the paper:** The figures 5a and 5b are now added to the revised version:



(a) **(b)**

Fig. 5. At *t*=1200 sec., comparison between the droplet size distributions obtained from the analytical solution of the master equation (line) and a) the SSA of Gillespie for 10^3 realizations (circles), b) the numerical algorithm (circles). Calculations were performed with the initial condition $P(30,0,0,0,...,0;0)=1$ for the sum kernel $K(i,j)=B(x_i+x_j)$, with $B=8.82\times10^2$ cm^3 sec^{-1}.

7. **References:**

This reference was missing and now is included in the revised version:

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

Additional references:

Bellman, R. (1961), Adaptive Control Processes: A Guided Tour, Princeton University Press.

Matsoukas, Themis. "Statistical Thermodynamics of Irreversible Aggregation: The Sol-Gel Transition." Scientific reports 5 (2015).

Nicolis, G., and Prigogine, L.: Self-Organization in Non-Equilibrium systems, Willey, N.Y., 1977.

Deleted references:

Gillespie, D. T.: The stochastic coalescence model for cloud droplet growth, *Journal of the Atmospheric Sciences*, *29*(8), 1496-1510, 1972.
Gillespie, D. T.: Three models for the coalescence growth of cloud drops, *Journal of the Atmospheric Sciences*, *32*(3), 600-607, 1975.

iii) A marked-up manuscript version showing the changes made in the manuscript (in the next page, with the additional text in blue).

An algorithm for the numerical solution of the multivariate master equation for stochastic coalescence.

L. Alfonso

[1]Universidad Autónoma de la Ciudad de México, México City, 09790 México

Correspondence to: Lester Alfonso (lesterson@yahoo.com)

Abstract

In cloud modeling studies, the time evolution of droplet size distributions due to collision-coalescence events is usually modeled with the Smoluchowski coagulation equation, also known as the kinetic collection equation (KCE). However, the KCE is a deterministic equation with no stochastic fluctuations or correlations. Therefore, the full stochastic description of cloud droplet growth in a coalescing system must be obtained from the solution of the multivariate master equation, which models the evolution of the state vector for the number of droplets of a given mass. Unfortunately, due to its complexity, only limited results were obtained for certain type of kernels and monodisperse initial conditions. In this work, a novel numerical algorithm for the solution of the multivariate master equation for stochastic coalescence that works for any type of kernels, multivariate initial conditions and small system sizes is introduced. The performance of the method was checked by comparing the numerically calculated particle mass spectrum with analytical solutions of the master equation obtained for the constant and sum kernels. Correlation coefficients were calculated for the turbulent hydrodynamic kernel, and true stochastic averages were compared with numerical solutions of the kinetic collection equation for that case. The results for collection kernels depending on droplet mass demonstrates that the

magnitude of correlations are significant, and must be taken into account when modeling the evolution of a finite volume coalescing system.

1. Introduction

The evolution of the size distribution of coalescing particles has often been described by the kinetic collection (hereafter KCE) or Smoluchowski coagulation equation, known under a number of names (“stochastic collection”, “coalescence”). The discrete form of this equation has the form (Pruppacher and Klett, 1997):

$$\frac{\partial N(i,t)}{\partial t}=\frac{1}{2}\sum_{j=1}^{i-1}K(i-j,j)N(i-j)N(j)-N(i)\sum_{j=1}^{\infty}K(i,j)N(j) \tag{1}$$

where $N(i,t)$ is the average number of droplets with mass x_i, and $K(i,j)$ is the collection kernel related to the probability of coalescence of two droplets of masses x_i and x_j. In Eq. (1), the time rate of change of the average number of droplets with mass x_i is determined as the difference between two terms: the first term describes the average rate of production of droplets of mass x_i due to coalescence between pairs of drops whose masses add up to mass x_i, and the second term describes the average rate of depletion of droplets with mass x_i due to their collisions and coalescence with other droplets.

Within the kinetic approach (Eq. (1), it is assumed that fluctuations are negligible small. This assumption can only be correct if the volume and the number of particles are infinitely large. An alternative approach considers the coalescence process in a system of finite number of particles, with fluctuations that are no longer negligible. This finite-volume description is intrinsically stochastic and has been pioneered by Marcus(1968), Bayewitz et al. (1974) and studied in detailed by Lushnikov (1978, 2004) and Tanaka and Nakazawa (1993).

Within the finite volume description a system of particles whose total mass is M_T is considered. The mass distribution of the particles is described by giving the number n_i of particles with mass i, i.e. n_1, n_2, $n_3, \ldots, n_N$. Then, the state of the mass distribution of the particle system is described by the N dimensional state vector $\bar{n} = (n_1, n_2, ..., n_N)$. The time evolution of the joint probability $P(n_1, n_2, ..., n_N; t)$ that the system is in state $\bar{n} = (n_1, n_2, ..., n_N)$ at time t is calculated according to the equation (Tanaka and Nakazawa, 1993):

$$\frac{\partial P(\bar{n})}{\partial t} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} K(i,j)(n_i+1)(n_j+1)P(..., n_i+1, ..., n_j+1, ..., n_{i+j}-1, ...; t)$$

$$+\sum_{i=1}^{N} \frac{1}{2} K(i,i)(n_i+2)(n_i+1)P(..., n_i+2, ..., n_{2i}-1, ...; t)$$

$$-\sum_{i=1}^{N} \sum_{j=i+1}^{N} K(i,j) n_i n_j P(\bar{n}; t) - \sum_{i=1}^{N} \frac{1}{2} K(i,i) n_i (n_i-1) P(\bar{n}; t) \qquad (2)$$

The master equation (2) is a gain-loss equation for the probability of each state $\bar{n} = (n_1, n_2, ..., n_N)$. The sum of the first two terms is the gain due to transition from other states, and the sum of the last two terms is the loss due to transitions into other states. The gain terms show that the system may be reached from any state with an i-mer and a j-mer more, and one $(i+j)$-mer less. In Eq. (2) $K(i, j)$ is the collection kernel and the transition rates are $K(i, j)(n_i+1)(n_j+1)$ if $i \neq j$ and $K(i,i)(n_i+1)(n_i+2)$ if $i = j$. From conservation of the total probability, $P(\bar{n}; t)$ must satisfy the relation:

$$\sum_{\bar{n}} P(\bar{n}; t) = 1 \qquad (3)$$

where the sum is taken over all states. Moreover, the total mass M_T of the system must be conserved, and the particle number n_i should be non-negative for any mass x_i:

$$\sum_{i=1}^{N} x_i n_i = M_T, \quad n_i \geq 0, \qquad i = 1, \ldots, N \tag{4}$$

Exact solutions of Eq. (2) are only known for a limited number of cases (constant, sum and product kernels) and for monodisperse initial conditions. For these special cases the master equation has been solved by Lushnikov (1978, 2004) and Tanaka and Nakazawa (1993) in terms of the generating function of $P(\bar{n};t)$. For general, multidisperse initial conditions, the solution of Eq. (2) is not known.

Additionally, for stochastic coagulation, approximate solutions were calculated by using the Van Kampen's system size expansion or Ω-expansion (Van Dongen and Ernst, 1987; Van Dongen, 1987) which permits to find solutions of Eq. (2) valid in the limit of a large system. However, the system size expansion gives less reliable results when applied to systems with a low number of particles or small volumes.

Then, in order to obtain solutions for more realistic kernels (Brownian motion, differential sedimentation etc.), a small number of particles and general multidisperse initial conditions, it has to be solved numerically. In this paper, we present an algorithm that can be applied to obtain the solution of Eq. (2) for any type of kernels and initial conditions. By applying this method, numerical solutions of the master equation were obtained for realistic kernels relevant to cloud physics, along with calculation of the correlations for the number of droplets for different sizes.

It is noteworthy to mention that the Stochastic Simulation Algorithm (SSA) developed by Gillespie (1975) also accurately reproduces the master equation. In Gillespie's method, the master equation is not solved directly, but a statistically correct trajectory (possible solution) of the master equation is generated. At any time, expected values at each droplet size can be obtained by averaging over many runs. However, a large number of realizations are necessary in order to obtain the desired accuracy at the large end of the droplet size distribution. A detailed comparison between the two methods will be made in section 3.

The problem of calculation of correlation coefficients was also addressed by Wang et al. (2006), who derived which they called the "true stochastic collection equation" (TSCE), which is a mean field equation at the first order and contains correlations among instantaneous droplets of different sizes. The problem with this equation and similar is that the rate change of moments of order n depends on moments of order $(n+1)$, as was remarked by Marcus (1968).

In our work, we overcome this drawback by calculating the true stochastic averages directly from the solution of the master equation. The main idea is to reduce the dimensionality by restricting the state space only to those states which have a finite probability of being accessed. It turns out that this provides a considerable improvement in numerical efficiency.

The paper is organized as follow: In section 2, the numerical algorithm is explained in detail. Numerical solutions for the sum and constant kernels with a comparison with analytical solutions and with *the method of Gillespie (1975)* are presented in section 3. The numerical results for mass dependent kernels along with calculation of correlations for different droplet sizes are presented in section 4. Finally, in section 5 we briefly discuss the results and the possible applications of the numerical algorithm.

2. The numerical algorithm

To solve Eq. (2) by brute force, the joint probability $P(n_1, n_2, ..., n_N; t)$ must be discretized into a multidimensional array. The main drawback of this approach is its susceptibility to the curse of dimensionality (Bellman, 1961), i.e. the exponential growth in memory and computational requirements in the number of problem dimensions.

For example, for a system with a mono-disperse initial condition $P(50,0,0,...,0;0)=1$, even considering the restriction (4), we would be in need to define a 50 dimensional array with about 1.34x10[16] elements, which is computationally prohibitive.

2.1 Calculation of all possible states.

Instead of the brute force discretization of the multi-dimensional joint probability distribution, the solution for this problem lies on the generation of all possible states from an initial configuration, and the posterior calculation of the time evolution of the probability $P(\bar{n};t)$ for each generated configuration by using the master equation. From an arbitrary initial condition $P(n_{01},n_{02},...,n_{0N};0)=1$ all possible states can be generated numerically. This can be performed by taking into account that the only allowed transitions are of the form: $\bar{n}_1^{(+)} \rightarrow \bar{n}_1$ if $i \neq j$ and $\bar{n}_2^{(+)} \rightarrow \bar{n}_2$ if $i = j$, where $\bar{n}_1^{(+)}, \bar{n}_1$ and $\bar{n}_2^{(+)}$, $\bar{n}_2$ are the state vectors:

$$\bar{n}_1^{(+)} = (n_1,...,n_i+1,...,n_j+1,...,n_{i+j}-1,...,n_N) \tag{5a}$$

$$\bar{n}_1 = (n_1,...,n_i,...,n_j,...,n_{i+j},...,n_N) \tag{5b}$$

$$\bar{n}_2^{(+)} = (n_1,...,n_i+2,...,n_{2i}-1,...,n_N) \tag{5c}$$

$$\bar{n}_2 = (n_1,...,n_i,...,n_{2i},...,n_N) \tag{5d}$$

For a system consisting of N monomers at $t=0$, $R(N)$ states (or N-dimensional vectors) can be realized, where $R(N)$ is the number of solutions in integers $\bar{n}$ of the Eq. (4) for conservation of mass. The number of possible configurations can be approximated from the equation (Hall, 1967):

$$R(N) \approx \frac{1}{4N\sqrt{3}} \exp\left(\pi\left(2N/3\right)^{1/2}\right) \tag{6}$$

Note that, although *R(N)* increases very quickly with *N* (for example, *R(*50*)*=217590 and *R(*100*)*=190 569 232), a number of states that is manageable with an average computer is obtained (compare with the 50 dimensional array with $1.34\text{x}10^{16}$ elements required for *N*=50). Although the formula (6) slightly overestimates the number of states, it gives estimates that can be used in order to check the performance of the algorithm. For *N*=6, 10, 20, 30 we obtained 11, 42, 627, and 5604 with the numerical algorithm, and 13, 48, 692 and 6078 by using the formula (6). As an example, the 11 possible configurations generated from the initial state $(6,0,0,0,0,0)$ are displayed in Fig. 1.

2.2 Time evolution of the probabilities $P(\overline{n};t)$ *for each state.*

At t_0=0, for the initial state $P(n_{01},n_{02},n_{03},n_{04},...,;t_0)=1$, and the probabilities for the rest of the states are set equal to 0. The probabilities of all generated configurations are updated according to the first order finite difference scheme:

$$\begin{aligned}
P\left(\overline{n};t_0+\Delta t\right) = P\left(\overline{n};t_0\right) \\
&+\Delta t\sum_{i=1}^{N}\sum_{j=i+1}^{N} K(i,j)(n_i+1)(n_j+1) \\
&\times P(...,n_i+1,...,n_j+1,...,n_{i+j}-1,...;t_0) \\
&+\Delta t\sum_{i=1}^{N}\frac{1}{2}K(i,i)(n_i+2)(n_i+1) \\
&\times P(...,n_i+2,...,n_{2i}-1,...;t_0) \\
&-\Delta t\sum_{i,j=1}^{N}K(i,j)n_i n_j P(\overline{n};t_0) \\
&-\Delta t\sum_{i=1}^{N}\frac{1}{2}K(i,i)n_i(n_i-1)P(\overline{n};t_0)
\end{aligned} \tag{7}$$

It is clear from Eq. (7) that the state probabilities $P\left(\overline{n};t_0+\Delta t\right)$ at $t=t_0+\Delta t$ will increase if the states from which transitions are allowed, have a non-zero probability at $t = t_0$ (second and third terms in the right-hand size of Eq. (6)), and will decrease due to collisions of particles from the same state at $t = t_0$ (fourth term and fifth terms in the right-hand side of Eq. (7)) if $P(\overline{n};t_0)$ is positive. The finite difference equation for $P(1,0,0,0,1,0)$ was written to illustrate the method. As can be checked from the generation scheme displayed in Figure 1, the only allowed transitions to $(1,0,0,0,1,0)$ are from the states $(1,1,1,0,0,0)$ and $(2,0,0,1,0,0)$. Consequently, at $t=t_0+\Delta t$, $P(0,1,0,1,0,0;t_0+\Delta t)$ will increase if $P(1,1,1,0,0,0;t_0)$ and $P(2,0,0,1,0,0;t_0)$ are positive at $t = t_0$. On the other hand, $P(1,0,0,0,1,0;t_0+\Delta t)$ will decrease due to collisions from particles within the same state at $t = t_0$ if $P(1,0,0,0,1,0;t_0)$ is positive. Then, $P(1,0,0,0,1,0;t_0+\Delta t)$ is calculated from the equation:

$$
\begin{aligned}
P\left(1,0,0,0,0,1,0;t_0+\Delta t\right) = P\left(1,0,0,0,0,1,0;t_0\right) \\
+\Delta t K(2,3)(n_2+1)(n_3+1)P(1,1,1,0,0,0;t_0) \\
+\Delta t K(1,4)(n_1+1)(n_4+1)P(2,0,0,1,0,0;t_0) \\
-\Delta t K(1,5)(n_1)(n_5)P(1,0,0,0,1,0;t_0)
\end{aligned}
\tag{8}
$$

In the second term of the right-hand side of (8), n_2+1 and n_3+1 are set equal to 1, as they are the number of particles in the second and third bins in the configuration $(1,1,1,0,0,0)$ at $t = t_0$. In the third term, $n_1+1=2$ and $n_4+1=1$ as they are defined from the state $(2,0,0,1,0,0)$, and finally $n_1=n_5=1$ in the fourth and last term. As an exercise, the time evolution of each state probability was calculated for the coalescence kernel $K(i,j)=(i^{1/2}+j^{1/2})/40$ from Marcus (1968). The results for 5 of the 11 possible configurations are displayed in Fig. 2.

2.3. Calculation of the expected values of the number of particles for each particle mass.

The number of particles for a given mass $n_1, n_2, \ldots, n_N$ are discrete random variables whose probability distributions can be obtained from:

$$P(n,m;t) = \sum_{Except\ \ n_m} P\left(n_1, n_2, \ldots, n_m = n, \ldots n_N; t\right) \tag{9}$$

Usually, the numerical implementation of Eq. (9) would involve calculating the sum of all elements of a multidimensional array, which is computationally very expensive. Our approach is simpler: Once the probabilities of all possible states are determined for all times, $P(n,m;t)$ can be calculated just by summing over all states that have n_m=n:

$$P(n,m;t) = \sum_{All\quad states\quad with\quad n_m = n} P\left(n_1, n_2, \ldots, n_m = n, \ldots n_N; t\right) \tag{10}$$

The expected values $\langle n_m \rangle$ for the number of particles of mass m are then calculated from the equation:

$$\langle n_m \rangle = \sum_n nP(n,m;t) \tag{11}$$

As an example, for the system from Fig. 1, the probability distribution $P(n,1;t)$ of having n particles with mass m=1 is displayed in Table 1.

3. Comparison with analytical solutions and the Stochastic Simulation Algorithm (SSA) of Gillespie.

3.1 Comparison with analytical solutions

The expected values for each particle mass calculated with the numerical algorithm, were tested against the analytical solutions of the master equation reported in Tanaka and Nakazawa (1993) for the constant (Eq. (12)) and sum (Eq. (13)) kernels ($K(i,j) = A$,

$K(i,j)=B(x_i+x_j)$) obtained for the mono-disperse initial condition $P(N_0,0,0,...,0;0)=1$. They are:

$$\langle n_m\rangle = C_m^{N_0} m! \sum_{l=1}^{N_0-m+1}\sum_{k=l}^{N_0}(-1)^{k-1}\frac{(2k-1)C_{l-1}^{N_0-1}C_{l-1}^{N_0-m}C_{N_0-k}^{N_0-l}}{(k+l-1)C_{k+l-1}^{N_0+k-1}} \times\left\{(l-1)\Big/\prod_{i=1}^{m}(N_0-i)\right\}e^{\frac{-k(k-1)}{2}\tau} \tag{12}$$

$$\langle n_m\rangle = C_m^{N_0}\left(\frac{i}{N_0}\right)^{m-1}\left\{1-\frac{m}{N_0}\left(1-e^{T}\right)\right\}^{N_0-m-1} \times\left(1-e^{-T}\right)^{m-1}e^{-T} \tag{13}$$

In Eqs. (12) and (13), N_0 is the initial number of particles, $C_m^{N_0}$ is the binomial coefficient and $\langle n_m\rangle$ are the true stochastic averages for each particle mass *m* at time *t*. In (12) $\tau=AN_0t$, where $A=1.2\times10^{-4}$ cm^3 sec^{-1} is the constant collection kernel. Finally, in (13), $T=BN_0v_0t$ where v_0 is the initial volume of droplets and $B=8.82\times10^{2}$ cm^3 g^{-1} sec^{-1}. Turning to a concrete numerical example, the evolution of a cloud system with an initial monodisperse droplet size distribution of $N_0=10$ droplets of 10 μm in radius (droplet mass 4.189×10^{-9}g) at t_0 , and a volume of 1cm^3 was calculated with the numerical algorithm. The time step was set equal to $\Delta t=0.1$ sec. A comparison between the numerical and analytical results for both the sum and constant kernels at $t=1200$ are shown in Figs. 3 and 4 with an excellent agreement between the two approaches.

3.2 Comparison with the SSA of Gillespie.

As was mentioned in the introduction, the algorithm of Gillespie generates a statistically correct trajectory of the stochastic master equation. It was presented in Gillespie (1975), and popularized in Gillespie (1977) were it was used to simulate chemical systems. As we know, in Gillespie's SSA, the ensemble mean for the number of droplets at each droplet mass is calculated from the expression (Gillespie, 1975):

$$N(m;t) = \frac{1}{N_r}\sum_{i=1}^{N_r} N^i(m;t) \tag{14}$$

where N_r is the number of realizations of the stochastic algorithm, $N^i(m;t)$ is the number of droplets of mass m in the i-realization at time t, and $N(m;t)$ is the ensemble mean. From expression (14**)** it is clear that in order to obtain the correct expected values $(N(m;t))$ at the large end of the droplet size distribution; we will need a large number of realizations of the SSA.

To further investigate this question, the evolution of a cloud system with an initial mono-disperse droplet size distribution of N_0=30 droplets of 14 μm in radius (droplet mass 1.1494×10^{-8}g) at t_0, and a volume of 1cm^3 was calculated with both the numerical algorithm and the Gillespie's SSA for the sum kernel ($K(i,j) = B(x_i + x_j)$, with B=8.82×10^2 cm^3 sec^{-1}) . The results obtained by the two methods, were then compared with the analytical solution of the master equation (Eq. (13)) obtained by Tanaka and Nakazawa (1993) for the same conditions.

The averages calculated from the Gillespie's method for N_r=10^3 realizations, and the analytical solution at t=1200 are displayed in Fig. 5. As can be observed, both the Monte Carlo averages and the analytical solution are closely coincident for the small end of the droplet size distribution. However, due to the small number of realizations, the SSA fails to reproduce the distribution for the expected values at the large end (See Table 1).

For a more detailed analysis, the expected number of particles for each droplet size calculated from the analytical solution, the numerical algorithm and the SSA of Gillespie (for 1000 and 10,000 realizations) are displayed in Table 1. As can be checked in the table, the size distributions are almost identical for the small end. However, they differ substantially at the large end since the SSA produces no particles larger than $12v_0$ and $16v_0$ for 1000 and 10,000 realizations respectively (v_0 =1.1494×10^{-8}g, mass of a 14 μm droplet). For 1000 realizations, the Monte-Carlo averages differ from the analytical solution for bin numbers larger than 8. For 10,000 realizations we have the same situation for bin numbers larger than 13.

As expected, for 1000 and 10000 realizations, no states with droplets 30 times larger than monomer-sized ones were realized. The numerical algorithm described in this paper performed very well at the large end, with expected values that are very close to the analytical solution (see Fig5 and Table 3).

It can be concluded that our method will be suitable if we need to accurately calculate the large end of the droplet spectrum for small systems (with <50 monomer droplets in the initial state). As the SSA requires a large number of realizations, it will be computationally very expensive. Then, for a small number of particles, our algorithm will be a good alternative, as it provides the desired accuracy to detect the possible small differences between different numerical approaches. It can also work as a benchmark for different Monte Carlo methods for the collision coalescence process.

4. Kinetic vs. stochastic approach: Calculation of correlation coefficients and numerical results for mass dependent collection kernels.

4.1 Numerical calculation of correlation coefficients

The evolution equation for the expected values of the random variables can be obtained by multiplying Eq. (7) by n_k and summing over all states (see Bayewitz et al. (1974)):

$$\frac{\partial}{\partial t}\langle n_k\rangle = \frac{1}{2}\sum_{i+j=k} K(i,j)\left(\langle n_i n_j\rangle - \langle n_i\rangle \delta_{i,j}\right) - \sum_j K(j,n)\left(\langle n_j n_k\rangle - \langle n_j\rangle \delta_{j,\mathrm{k}}\right) \quad (15)$$

The KCE is obtained from (15) by assuming that $\langle n_i n_j\rangle = \langle n_i\rangle\langle n_j\rangle$, i.e. that the correlation between the random variables is zero. A form of Eq. (15) was deduced in Tanaka and Nakazawa (1993) and in Wang et al. (2006) for a general type of kernels. Bayewitz et al. (1974) have quantified the deviation of the size distributions calculated with the KCE from the exact distribution obtained from the master equation for a constant kernel. From Eq. (15) it can be concluded that as long as the correlations remain appreciable, the results of the KCE will not match the true stochastic averages. The correlation (or correlation coefficient) between two random variables n_i and n_j denoted as $\rho_{i,j}$ is

$$\rho_{i,j} = \frac{\mathrm{cov}(n_i, n_j)}{\sqrt{Var(n_i)Var(n_j)}} = \frac{\sigma_{n_i n_j}}{\sigma_{n_i}\sigma_{n_j}} \quad (16)$$

In **(16),** the covariance ($\mathrm{cov}(n_i, n_j)$) is calculated according to

$$\mathrm{cov}(\mathrm{n}_i, \mathrm{n}_j) = E\left[\left(n_i - \langle n_i\rangle\right)\left(n_j - \langle n_j\rangle\right)\right] = E\left(n_i n_j\right) - \langle n_i\rangle\langle n_j\rangle \quad (17)$$

Where $E\left(n_i n_j\right)$ is the expected value of the product $n_i n_j$, which, for the bivariate case is:

$$E\left(n_i n_j\right) = \sum_{n_i}\sum_{n_j} n_i n_j f(n_i, n_j) \quad (18)$$

In Eq. (18), $f(n_i, n_j)$ is the two dimensional joint probability mass function (*pmf*) which was calculated similarly to how it was done in the univariate case (See Eq. 10):

$$f(n,l) = P(n,i;l,j;t) = \sum_{\substack{All \quad states \quad with \\ n_i = n \quad and \quad n_j = l}} P\left(n_1, n_2, ..., n_i = n, ..., n_j = l, ..., n_N; t\right) \tag{19}$$

In the former equation, $P(n,i;l,j;t)$ is the probability of having n droplets of mass i and l droplets of mass j.

4.2 Numerical results for the constant, sum and product kernels.

Correlation coefficients ($\rho_{1,2}$ and $\rho_{2,3}$) were obtained by Wang et al. (2006) using the analytical solution obtained by Bayewitz et al. (1974) for a constant collection kernel. They found that, even for this case, the magnitude of correlations could be quite large. We will extend their analysis by calculating the time evolution of the correlation coefficients $\rho_{1,2}$ and $\rho_{2,3}$ for the constant, sum and product kernels (see Fig. 6). For each case, the simulations were conducted for two systems containing 10 and 40 droplets of 14 μm in radius respectively, and a volume of 1cm^{-3}. As can be observed in the figure, in all the cases we have non-zero correlations. From the evolution of $\rho_{1,2}$ for all the kernels, we can infer that the random variables n_1 and n_2 are, at the beginning of the simulation, strongly anticorrelated. This is due to the fact that in the initial stage of evolution of the system we have mainly collisions between size 1 droplets to form size 2 droplets. On the other hand, the random variables n_2 and n_3 are also anticorrelated, because a decrease of n_2 due to collisions with size 1 droplets will increase the number of size 3 droplets (Wang et al., 2006).

At t=1800 sec., the true stochastic averages (see Eq. 11) obtained numerically from the master equation are displayed in Fig. 7, together with the mean values for each droplet mass calculated from the analytical solutions of the KCE (See Table 2.). For the three cases, at the large end of the spectrum, results differ substantially. This is in agreement with the analytical study of Tanaka and Nakazawa (1994), who demonstrated that the true stochastic averages coincide well with those obtained from the kinetic collection equation (1) if the bin mass k satisfies the inequality $k^2 \ll M_0$, where M_0 is the total mass of the system.

4.3 Numerical results for the turbulent hydrodynamic collection kernel.

Collisions between droplets under pure gravity conditions are simulated with a collection kernel of the form:

$$K_g(x_i, x_j) = \pi (r_i + r_j)^2 \left| V(x_i) - V(x_j) \right| E(r_i, r_j) \tag{20}$$

The hydrodynamic kernel (Eq. 19) doesn't take into account the turbulence effects and considers that droplets with different masses (x_i and x_j and corresponding radii, r_i and r_j) have different settling velocities. In Eq. 20, $E(x_i\ ,x_j)$ are the collection efficiencies calculated according to Hall (1980). In turbulent air, the hydrodynamic kernel should be enhanced due to an increase in relative velocity between droplets (transport effect) and an increase in the collision efficiency (the drop hydrodynamic interaction). These effects were taken into account by implementing the turbulence induced collision enhancement factor $P_{Turb}(x_i, x_j)$ calculated in Pinsky et al. (2008) for a cumulonimbus cloud with dissipation rate, ε=0.1 m^{2}s^{-3}and Reynolds number, Re_λ=2×10^4 for cloud droplets with radii ≤ 21 μm. Then, the turbulent collection kernel has the form:

$$K_{Turb}(x_i, x_j) = P_{Turb}(x_i, x_j) K_g(x_i, x_j) \tag{21}$$

In the simulation for turbulent air, a system corresponding to a cloud volume of 1cm^3 and a bidisperse droplet distribution was considered: 20 droplets of 14 μm in radius and another 10 droplets of 17.64 μm in radius, corresponding to a liquid water content (LWC) of 0.436 gm^{-3}. For the turbulent collection kernel the true stochastic averages at t=200, 1800 sec. are displayed in Fig.8, and compared with the mean values for each droplet mass calculated numerically from the kinetic collection equation (KCE) with kernel (20). Also, for this case, at the large end of the spectrum, results obtained from the KCE differ substantially from the stochastic means. The time evolution of the correlation coefficients $\rho_{1,5}$ and $\rho_{1,3}$ displayed in Fig. 9 confirms the fact that correlations cannot be neglected.

Finally, the time variation of $\langle n_1 \rangle$, $\langle n_5 \rangle$, $\langle n_{15} \rangle$ and $\langle n_{20} \rangle$ were calculated and compared with the time evolution of the averages calculated from the KCE with the same initial conditions and coalescence rate. We can see from Fig. 10 that for the small masses k=1 and 5, both solutions are closely coincident up to 1800 sec., and that for the larger masses k=15 and 20, the results are different at all times.

5. Discussion and conclusions.

The full stochastic description of the growth of cloud droplets in a coalescing system is a challenging problem. For finite volume systems or in systems of small populations, statistical fluctuations become important and the mathematical description relies on the master equation which has analytical solutions for a limited number of cases. In an effort to solve this problem, we have introduced a new approach to numerically calculate the solution of the coalescence multivariate master equation that works for any type of kernels and initial conditions.

For the constant, sum and product kernels, the true stochastic averages calculated numerically were compared with analytical solutions of the master equation, with an excellent agreement between the two approaches.

A numerical procedure to calculate the correlation coefficients was implemented, which were calculated for mass dependent kernels (sum, product, and kernels modified by turbulent processes). Also numerical solutions of the master equation for bivariate initial conditions and collection kernels modified by turbulent processes were obtained and compared with size distribution obtained from the numerical integration of the KCE. The two equations give different values at the large end of the droplet size distribution. It was also shown that, for small k, the true stochastic averages $\langle n_k \rangle$ and the solution of the KCE are closely coincident up to 1800 sec. For larger masses, the results are different at all times.

Can be a topic of discussion the limits of applicability of the finite volume approach to problems of precipitation formation, since such small volumes would not remain undisturbed for a long time in a real cloud. However, in defense of the finite system approach, it might be argued that, in the early stages of cloud development, due to small terminal velocities of the droplets, the coalescence process is a fairly localized process. Then, two droplets in widely separated parts of the cloud are not going to be coalescing with each other. This was the approach followed by Bayewitz et al. (1974) (and endorsed in Gillespie (1975)). In their paper, for comparing the stochastic and kinetic approaches, they partitioned the cloud into many sub-volumes, with no collisions being permitted for two droplets of different sub-volumes. However, interactions between sub-volumes through sedimentation, diffusion or other physical processes were not considered.

For a constant collection kernel, a more complex model that uses the master equation formalism, and introduces the interactions between the sub-volumes was developed by Merkulovich and Stepanov (1990, 1991). This model is based on a scheme proposed by

Nicolis and Prigogine (1977) for chemical reactions. Within this theory, the whole system is subdivided into sub-volumes (coalescence cells) that can be considered spatially homogeneous. Coalescence events are permitted only between droplets from the same sub-volume, and interactions between neighbors occur through the diffusion process. That leads to a set of master equations for each sub-volume. Although very complex, it could be a starting point in order to consider the interactions between small coalescence volumes through sedimentation or other physical mechanisms.

However, fluctuations will be also very important, if the collection kernel K(i,j) increases sufficiently rapidly with i and j and a giant droplet with mass comparable to the total mass of the system is formed. In that case, the total mass predicted by the KCE starts to decrease. This is usually interpreted to mean that the system exhibits a phase transition (also called gelation). After this moment, the true averages calculated from the master equation will differ from the averages obtained from Eq. 1, and there is a transition from a system with a continuous droplet distribution to one with a continuous distribution plus a giant cluster (Alfonso et al., 2013). After the sol-gel transition the KCE breaks down: the second moment of the size distribution diverges at the gel point, and, as was remarked, the first moment decays, i.e., mass is not conserved.

The limitation of the KCE equation arises from the fact that it is a deterministic equation with no fluctuations or correlations included. Then it describes an inherently stochastic process with a single metric, the mean cluster distribution (Matsoukas, 2015). Then, in order to model properly the system behavior after the giant cluster is formed, the role of fluctuations should be considered.

By using the finite volume approach, the expected values at the large end of the droplet size distribution can be obtained in the post-gel region (Lushnikov, 2004; Matsoukas, 2015), and be compared with the expected values obtained from the kinetic approach. As a result, it is expected to obtain broader droplet mass distributions by using the stochastic approach. A follow-up paper will be devoted to a more detailed analysis of all these problems.

Acknowledgements. The author thanks the Associate Program of the Abdus Salam International Center of Theoretical Physics (ICTP), in Trieste, Italy, where this work was started in the summer of 2013. This study was funded by grants UACM-SECITI No. PI2011-68R and Consejo Nacional de Ciencia y Tecnología de México SEP-CONACYT CB-131879.

References

Alfonso, L., Raga, G.B., Baumgardner, D.: The validity of the kinetic collection equation revisited. *Atmos. Chem. Phys*., 8, 969-982, 2008.

Alfonso, L., Raga, G.B., Baumgardner, D.: The validity of the kinetic collection equation revisited. Part II: Simulations for the hydrodynamic kernel, *Atmos. Chem. Phys*., 10, 6219-6240, 2010.

Alfonso, L., Raga, G. B., 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.

Bellman, R.E.: Adaptive control processes: A guided tour. Princeton University Press (Princeton, NJ), 1961.

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.

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.: Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, *81*(25), 2340-2361, 1977.

Hall, M.J.: Combinatorial Theory. Blaisdell Pub. Co., 1967.

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

Lushnikov, A. A.: From sol to gel exactly. *Physical review letters*, vol. 93, no 19, p. 198302, 2004,

Lushnikov, A. A.: Coagulation in finite systems. *Journal of Colloid and Interface Science*, vol. 65, no 2, p. 276- 285, 1978.

Marcus, A. H.: Stochastic coalescence, *Technometrics*, 10.1, 133-143, 1968.

Malyshkin, L., and Goodman, J.: The timescale of runaway stochastic coagulation, *Icarus*, *150*(2), 314-322, 2001.

Nicolis, G., and Prigogine, L.: Self-Organization in Non-Equilibrium systems, Willey, N.Y., 1977.

Matsoukas, Themis. "Statistical Thermodynamics of Irreversible Aggregation: The Sol-Gel Transition." Scientific reports 5 (2015).

Pruppacher, H.R., Klett, J.D.: Microphysics of clouds and precipitation, Kluwer Academic Publishers, 1997.

Richtmeyer, D., and Morton, K.W.: *Difference Methods for Initial Value Problems*, 2nd ed., Wiley, New York, 1967.

Tanaka, H., Nakazawa, K.: Stochastic coagulation equation and the validity of the statistical coagulation equation, *J. Geomag. Geoelecr*., 45, 361-381, 1993.

Van Dongen, P. G. J., and Ernst, M. H.: Fluctuations in coagulating systems, *Journal of statistical physics*, vol. 49, No 5-6, p. 879-926, 1987.

Van Dongen, P. G. J.: Fluctuations in coagulating systems. II. *Journal of statistical physics*, vol. 49, no 5-6, p. 927-975, 1987.

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

Wetherill, G.W.: Comparison of analytical and physical modeling of planetesimal Accumulation, *Icarus* 88, 336–354, 1990.

Table 1. Probability distribution $P(n,1;t)$ of finding n particles of size $m=1$ at time t, for a system with the initial condition $P(6,0,0,0,0,0;0)=1$.

	Probability distribution $P(n,1;t)$
n=0	$P(0,1;t)=P(0,1,0,1,0,0,t)+P(0,0,2,0,0,0)+P(0,0,0,0,0,1)$
n=1	$P(1,1;t)=P(1,1,1,0,0,0,t)+P(1,0,0,0,1,0)$
n=2	$P(2,1;t)=P(2,2,0,0,0,0;t)+P(2,0,0,0,1,0;t)$
n=3	$P(3,1;t)=P(3,0,1,0,0,0;t)$
n=4	$P(4,1;t)=P(4,1,0,0,0,0;t)$
n=5	$P(5,1;t)=0$
n=6	$P(6,1;t)=P(6,0,0,0,0,0;t)$

Table 2. Analytical size distributions of the kinetic collection equation (KCE) calculated with monodisperse initial conditions.

$K(x_i, x_j)$	$N(i,t)$	
$B(x_i + x_j)$	$N_0(1-\phi)\dfrac{(i\phi)^{i-1}}{\Gamma(i+1)}\exp(-i\phi)$	$\phi = 1 - \exp(-BN_0 v_0 t)$
$C(x_i \times x_j)$	$N_0 \dfrac{(iT)^{i-1}}{i\Gamma(i+1)}\exp(-iT)$	$T = CN_0 {v_0}^2 t$
A	$4N_0 \dfrac{(T)^{i-1}}{(T+2)^{i+1}}$	$T = AN_0 t$

Note: Parameters β, B and C are constants, x and y are the masses of the colliding drops. N_0 is the initial concentration and v_0 is the initial volume of droplets. The index i represents the bin size.

Table 3. Expected values for each droplet mass obtained at t=1200 sec. for the analytical solution, the numerical algorithm proposed in this work, and the Gillespie's SSA (for N_r= 1000, 10000 realizations). Calculations were performed with the initial condition $P(30,0,0,0,...,0;0) = 1$, and the sum kernel $K(i, j) = B(x_i + x_j)$ with B=8.82×10^2 cm^3 sec^{-1}.

Expected values for each droplet size: <n_i>, t=1200 sec.				
Bin Number	Analytical Solution	Numerical algorithm	SSA (N_r= 1000)	SSA (N_r= 10,000)
1.000	1.5633E+01	1.5622E+01	1.5612E+01	1.5619E+01
2.000	3.5302E+00	3.5303E+00	3.5250E+00	3.5425E+00
3.000	1.1754E+00	1.1762E+00	1.1870E+00	1.1712E+00
4.000	4.5543E-01	4.5609E-01	4.4800E-01	4.5050E-01
5.000	1.9017E-01	1.9057E-01	2.2300E-01	1.9600E-01
6.000	8.2592E-02	8.2824E-02	7.2000E-02	8.2000E-02
7.000	3.6583E-02	3.6709E-02	3.6000E-02	3.6800E-02
8.000	1.6320E-02	1.6387E-02	1.6000E-02	1.6100E-02
9.000	7.2696E-03	7.3034E-03	3.0000E-03	6.5000E-03
10.000	3.2117E-03	3.2284E-03	2.0000E-03	3.5000E-03
11.000	1.3997E-03	1.4077E-03	1.0000E-03	1.2000E-03
12.000	5.9891E-04	6.0263E-04	0.0000E+00	4.0000E-04
13.000	2.5049E-04	2.5216E-04	0.0000E+00	4.0000E-04
14.000	1.0197E-04	1.0269E-04	0.0000E+00	3.0000E-04
15.000	4.0229E-05	4.0529E-05	0.0000E+00	0.0000E+00
16.000	1.5312E-05	1.5431E-05	0.0000E+00	1.0000E-04
17.000	5.5954E-06	5.6404E-06	0.0000E+00	0.0000E+00
18.000	1.9526E-06	1.9687E-06	0.0000E+00	0.0000E+00
19.000	6.4672E-07	6.5217E-07	0.0000E+00	0.0000E+00
20.000	2.0189E-07	2.0361E-07	0.0000E+00	0.0000E+00
21.000	5.8917E-08	5.9419E-08	0.0000E+00	0.0000E+00
22.000	1.5913E-08	1.6048E-08	0.0000E+00	0.0000E+00
23.000	3.9295E-09	3.9622E-09	0.0000E+00	0.0000E+00
24.000	8.7349E-10	8.6634E-10	0.0000E+00	0.0000E+00
25.000	1.7127E-10	1.7176E-10	0.0000E+00	0.0000E+00
26.000	2.8809E-11	2.8765E-11	0.0000E+00	0.0000E+00
27.000	3.9922E-12	3.9906E-12	0.0000E+00	0.0000E+00
28.000	4.2746E-13	4.2803E-13	0.0000E+00	0.0000E+00
29.000	3.1450E-14	3.1525E-14	0.0000E+00	0.0000E+00
30.000	1.1930E-15	1.1962E-15	0.0000E+00	0.0000E+00

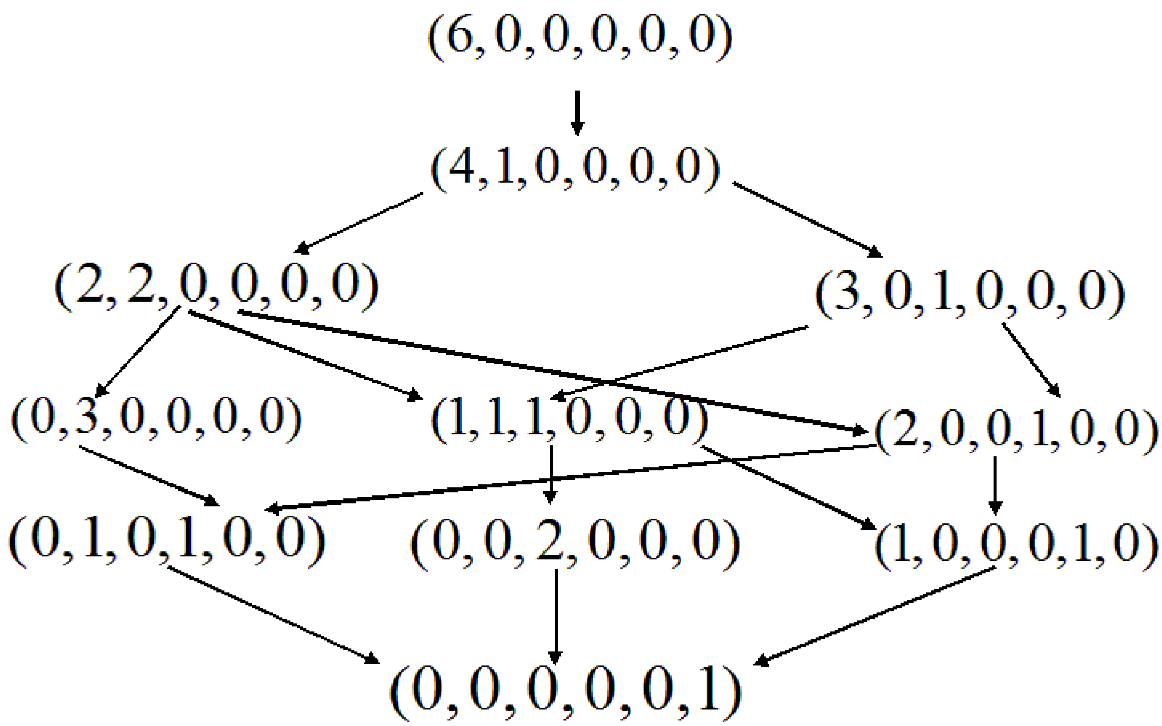


Fig. 1. States space obtained from the initial condition $P(6,0,0,0,0,0;0)=1$ with the constraint $\sum_{i=1}^{6} in_i = 6$.

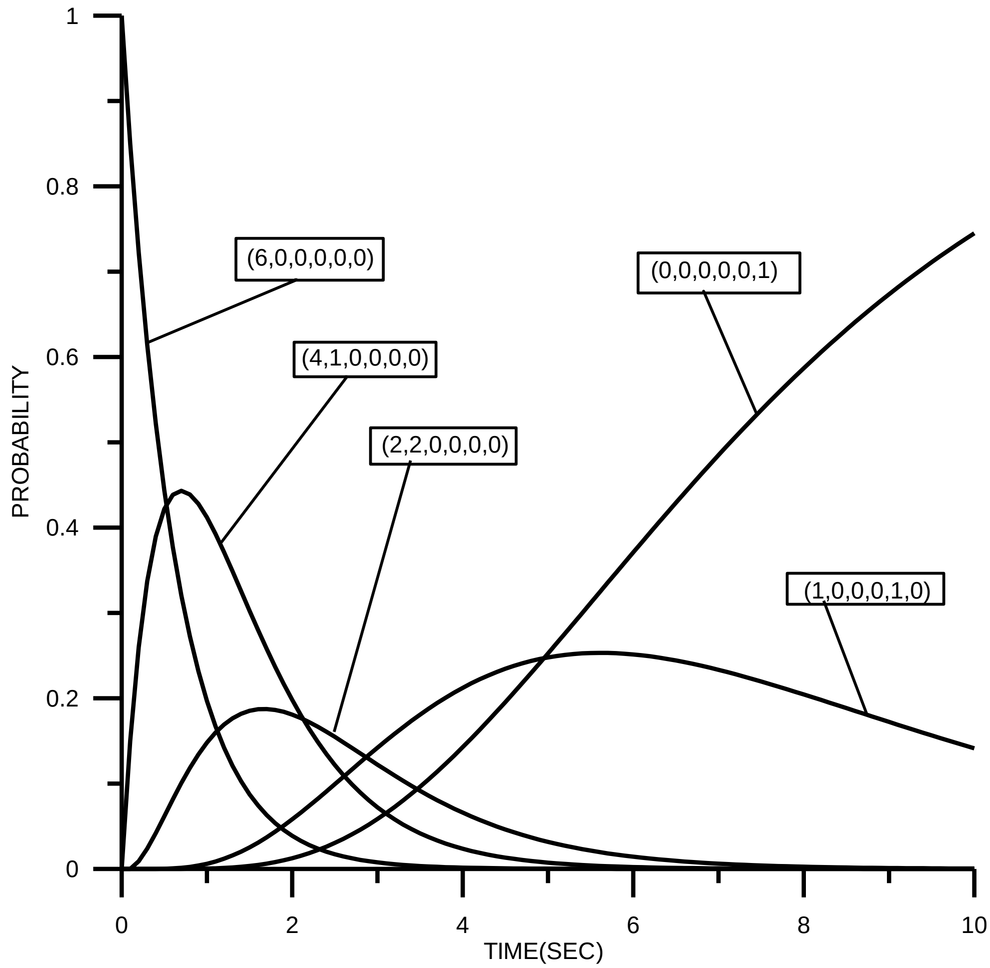


Fig. 2. Time evolution of the probability for 5 of the 11 the states for the initial condition $P(6,0,0,0,0,0;0)=1$ and the collection kernel $K(i,j)=(i^{1/2}+j^{1/2})/40$.

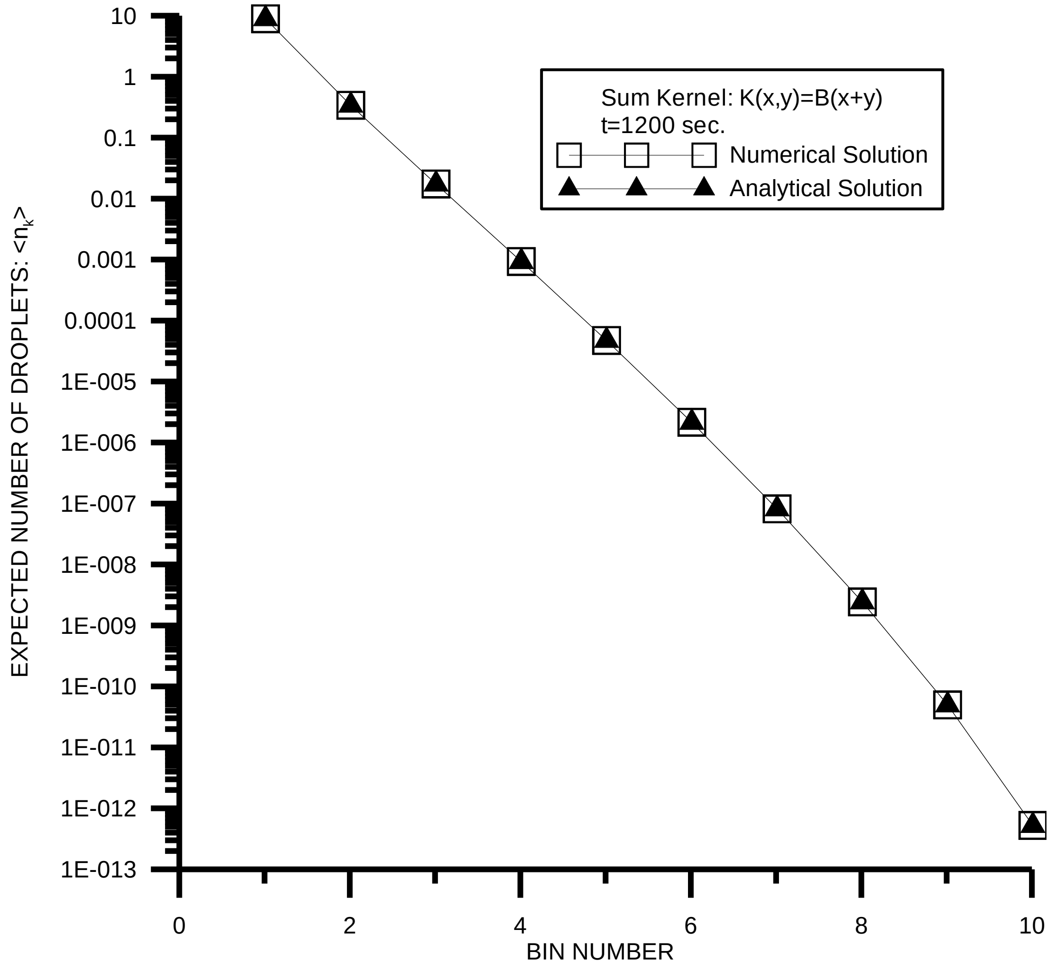


Fig. 3. For the sum kernel, size distribution obtained from the analytical solution of the master equation (triangles) and the numerical algorithm (squares) at t=1200 sec. Calculations were performed with the initial condition $P(10,0,0,...,0;0)=1$ and the sum kernel $K(i,j)=B(x_i+x_j)$, with B=8.82×10^2 cm^3 sec^{-1}.

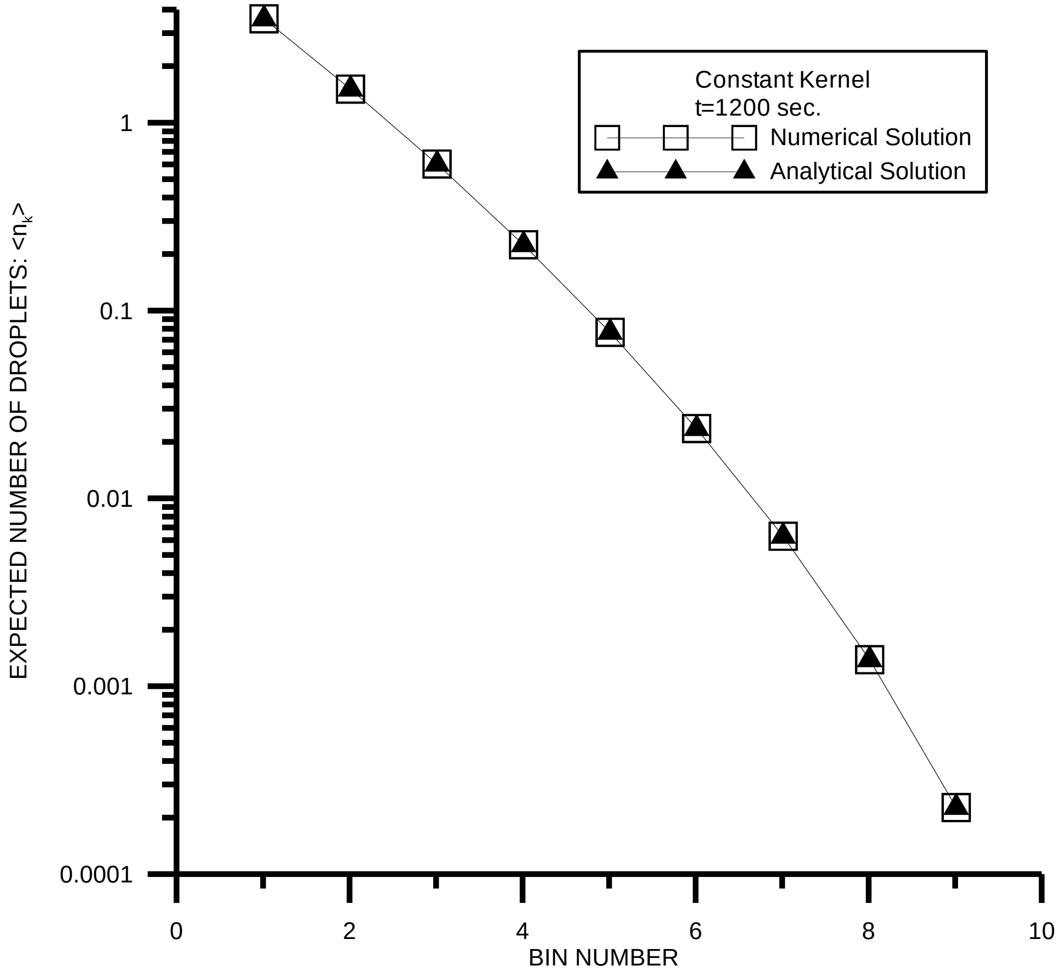
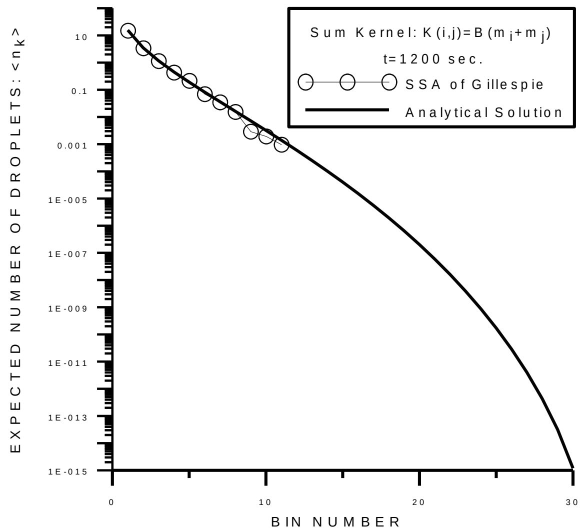
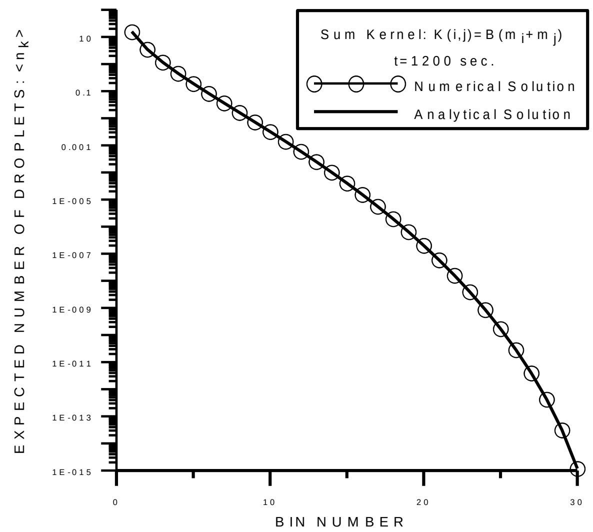


Fig. 4. Same as Fig. 3 but for the constant kernel $K(i, j) = 1.2\times10^{-4}\,\mathrm{cm}^{-3}$.

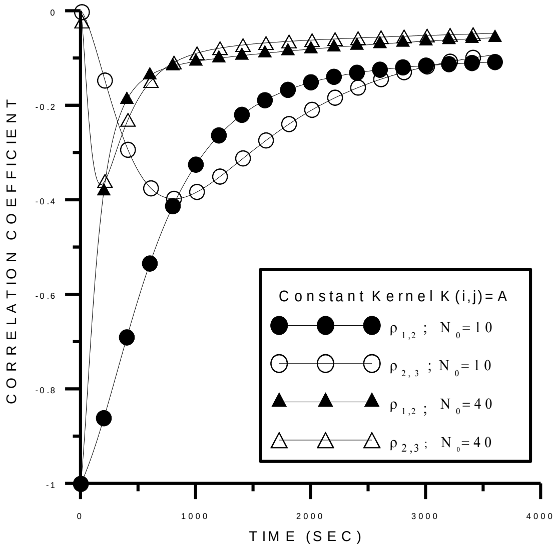


(b)

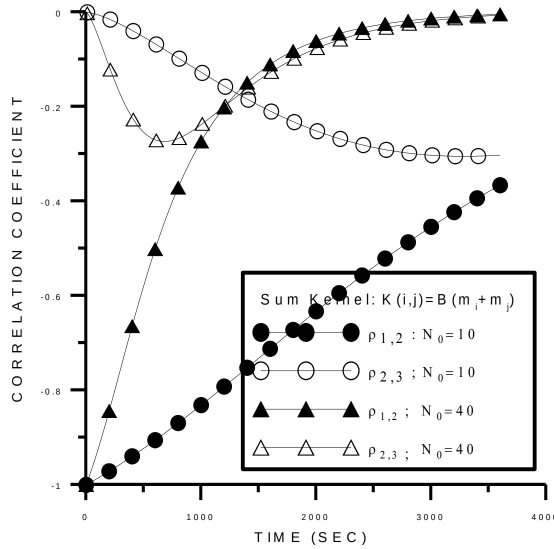


(b)

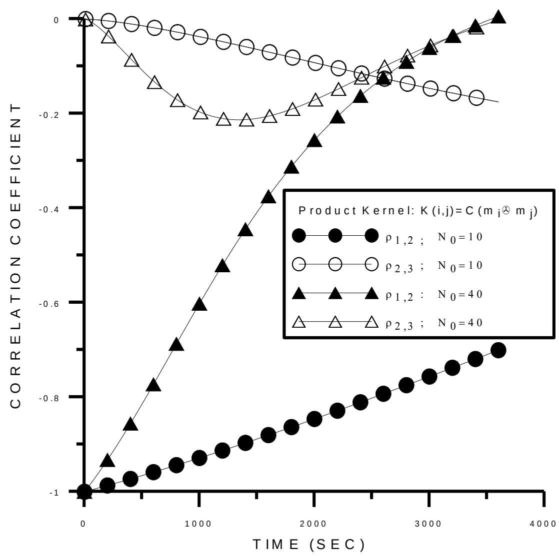
Fig. 5. At t=1200 sec., comparison between the droplet size distributions obtained from the analytical solution of the master equation (line) and a) the SSA of Gillespie for 10^3 realizations (circles), b) the numerical algorithm (circles). Calculations were performed with the initial condition $P(30,0,0,0,...,0;0)=1$ for the sum kernel $K(i,j)=B(x_i+x_j)$, with B=8.82×10^2 cm^3 sec^{-1}.



(a)

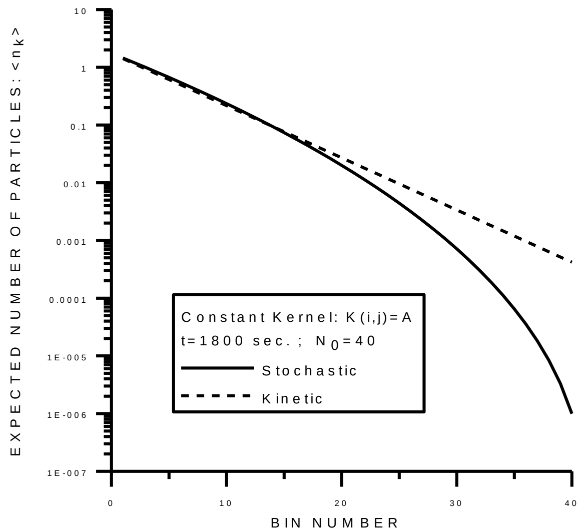


(b)

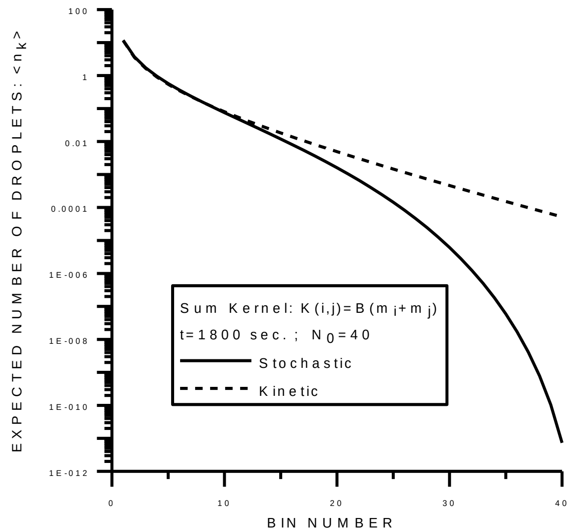


(c)

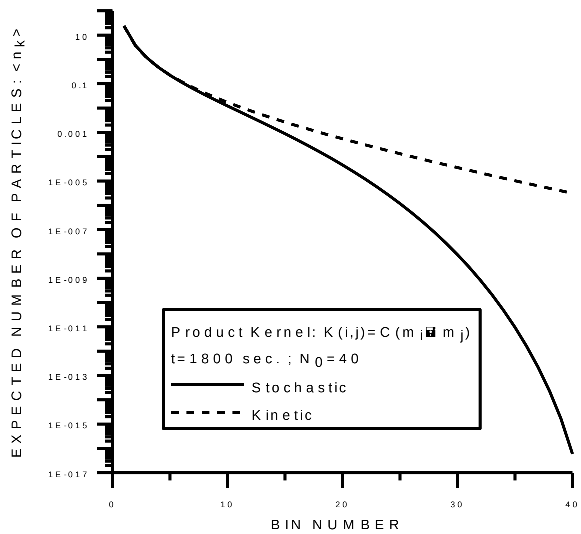
Fig. 6. Time evolution of the correlation coefficients $\rho_{1,2}$ and $\rho_{2,3}$ for the constant, sum and product kernels (in figures (a), (b) and (c) respectively) for two systems with a volume of 1 cm^{-3} and containing 10 and 40 droplets of 14 μm.



(a)

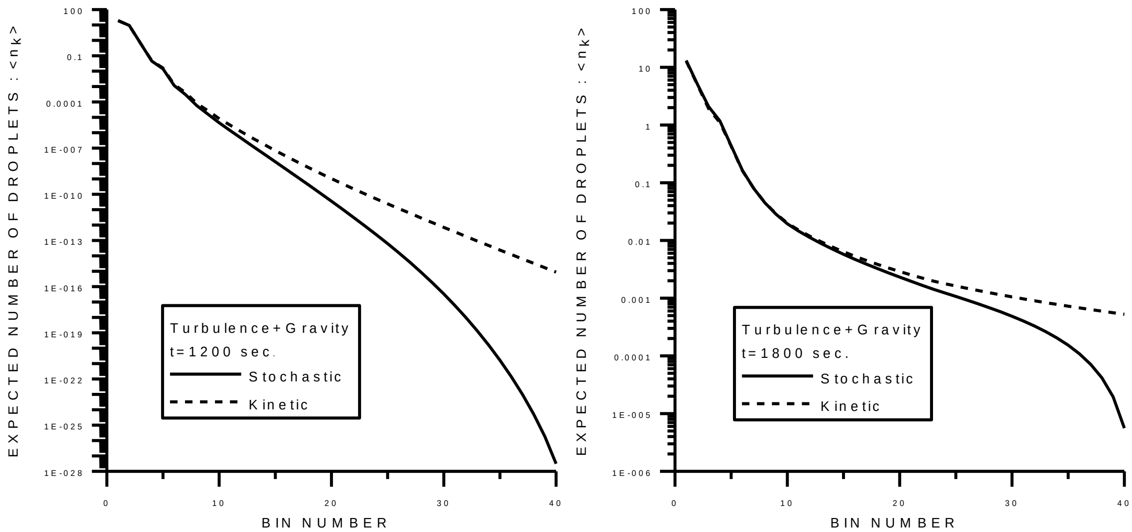


(b)



(c)

Fig. 7. Comparison of the size distributions obtained from the stochastic master equation (solid line) with that to the KCE (dashed line) at t=1800 sec. for a 1cm^3 system containing initially 40 droplets of 14 μm. The expectation values are shown for the constant, sum and product kernels (in figures (a), (b) and (c) respectively). For the small end the size distributions are closely coincident, for the large end the two equations give different values.



(a) (b)

Fig. 8. For the turbulent hydrodynamic kernel, comparison of the size distributions obtained from the stochastic master equation (solid line) with that to the KCE (dashed line) at *t*=200, 1800 sec. for a 1cm^3 system containing initially 20 droplets of 14 μm and 10 droplets of 17 μm. For the small end the size distributions are closely coincident, for the large end the two equations give different values.

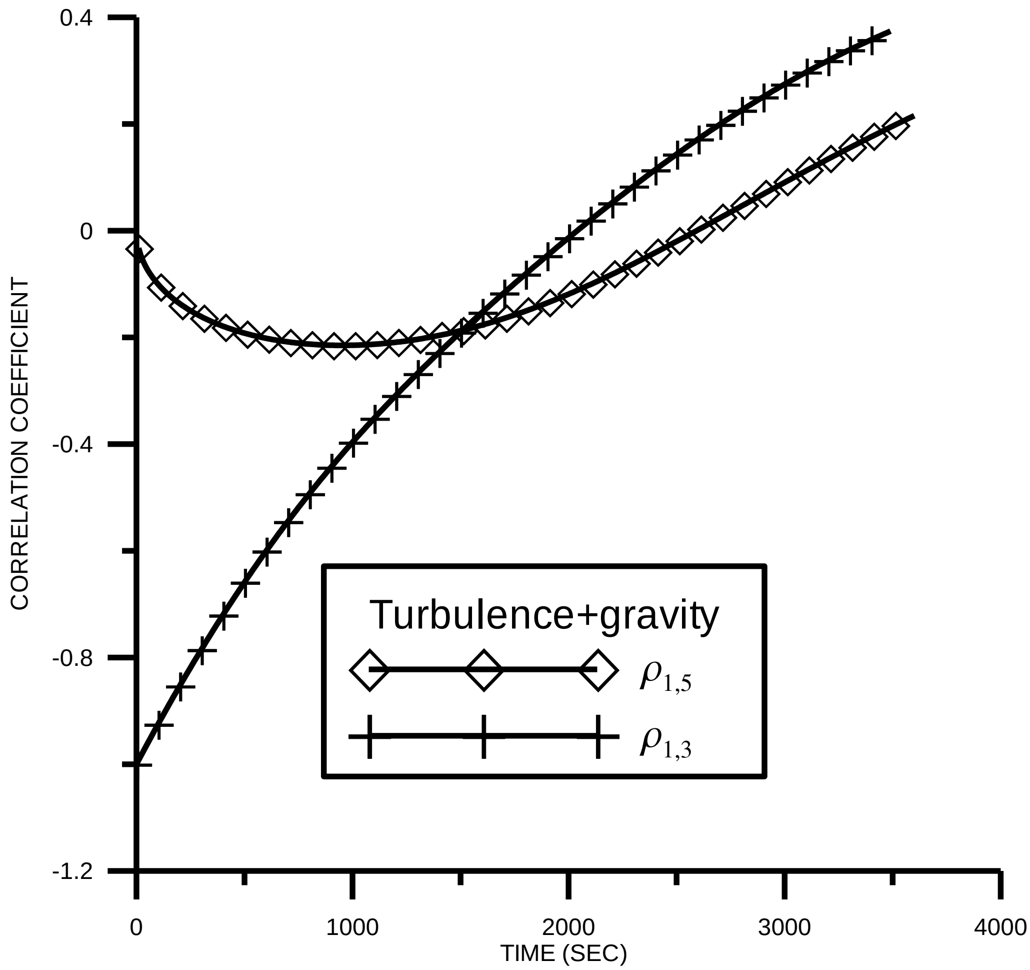
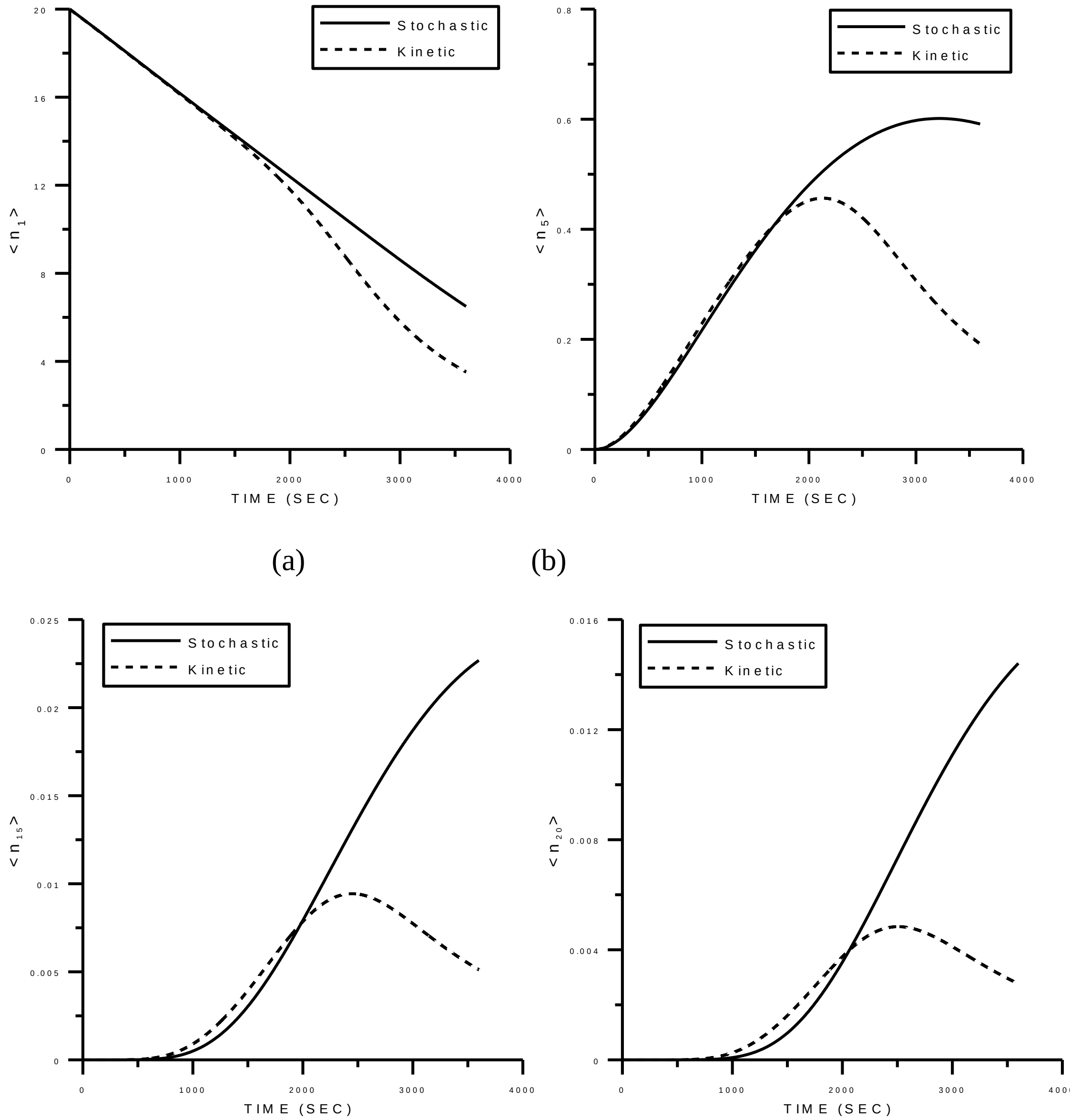


Fig. 9. Time evolution of the correlation coefficients $\rho_{1,3}$ and $\rho_{1,5}$ for a $1cm^3$ system modeled with the turbulent hydrodynamic kernel and containing initially 20 droplets of 14 μm and 10 droplets of 17 μm.



(a) (b)

(c) (d)

Fig. 10. For the turbulent hydrodynamic kernel, comparison of the expected values $\langle n_k \rangle$ obtained from the stochastic master equation (solid line) with that to the KCE (dashed line), for a 1cm^3 system containing initially 20 droplets of 14 μm and 10 droplets of 17 μm. The time evolution of the expected values are shown for k=1, 5, 15 and 20 (figures (a), (b), (c) and (d) respectively). For the small masses k=1 and 5, both solutions are closely coincident up to 1800 sec. For the larger masses k=15 and 20, the results are different at all times.