数理解析研究所講究録
第1606巻 2008年 110-118

Estimation of the Computational Cost of Super-Droplet Method

Shin-ichiro Shima, The Earth Simulator Center, Japan Agency for Marine-Earth Science and Technology

Abstract: Super-Droplet Method (SDM) is a novel, particle based, probabilistic simulation model of cloud microphysics. A theoretical analysis to evaluate the computational efficiency of SDM is developed. It is estimated that SDM will be computationally less demanding compared to the spectral (bin) method when the number of attributes, the internal degrees of freedom of each super-droplet, is increased and the accuracy of the spectral (bin) method is not so high-order.

1 Introduction

Although clouds play a crucial role in atmospheric phenomena, the numerical modeling of cloud is not yet well established. The fluid motion of the moist air associated with clouds is called "cloud dynamics process" and the behavior of aerosol/cloud/precipitation particles floating in the atmosphere is called "cloud microphysics process". These two processes mutually affect each other in the course of cloud formation and precipitation development, and this fact suggests that we need to simulate both processes and their interactions concurrently in order to produce an accurate prediction. Cloud dynamics model to describe the fluid motion in the atmosphere has been well developed [1]. However, it is still difficult to perform an accurate simulation of cloud microphysics though several simulation methods, such as bulk parameterization method [2–8], spectral (bin) method [9–13], and the exact Monte Carlo method [14, 15], have been proposed. Numerical methods to accurately simulate the cloud microphysics and the interactions between the cloud dynamics are required to understand and predict cloud-related phenomena.

Recently, the present author and his coworkers have developed a novel, particle based, probabilistic simulation model of cloud microphysics, named Super-Droplet Method (SDM) [16]. Though several extensions and validations are still necessary, we expect that SDM enables accurate numerical simulation of cloud microphysics with less demanding cost in computation. The methodology to couple super-droplets and the cloud dynamics model was also developed. We can use a compressible, non-hydrostatic model to simulate the cloud dynamics process (the fluid dynamical motion of the moist air). Figure 1 shows a snapshot of a shallow maritime cumulus formation simulated by the coupled model of SDM and a non-hydrostatic model. We can see that the turbulent like structures inside the cloud is resolved in our simulation. SDM provides us a new approach to the cloud-related open problems, such as the cloud and aerosol interactions, the cloud-related radiative processes, and the mechanism of thunderstorms and lightning.

The main objective of the present paper is the theoretical analysis of the computational cost of SDM. For this purpose, we focus our attention to simulate the stochastic coalescence process, the most computationally expensive elementary process of cloud microphysics process. Finally, it is estimated that SDM will be computationally less demanding compared to the spectral (bin) method when the number of attribute, the internal degrees of freedom of real-droplet, becomes larger and the accuracy of the spectral (bin) method is not so high-order.

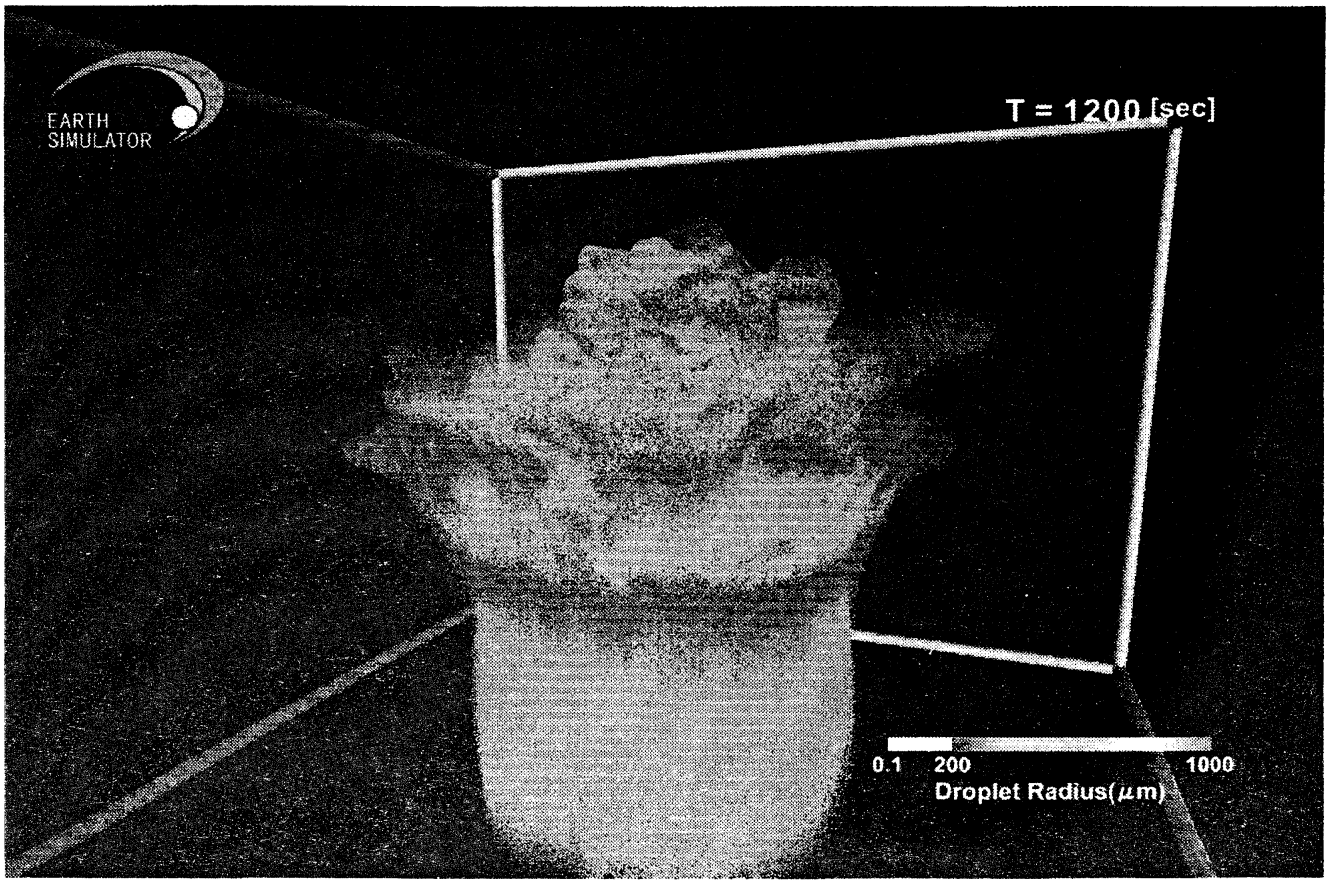

Figure 1: 3-dimensional simulation of a shallow maritime cumulus formation and precipitation using a coupled model of the Super-Droplet cloud microphysics model and a non-hydrostatic model.

2 Stochastic coalescence of real-droplets

Let us introduce a system in which the coalescence of real-droplets repeatedly occurs as a stochastic event.

We use the word *real-droplet* as a generic term referring to the aerosol/cloud/precipitation particles. Let $N_r(t)$ be the number of real-droplets floating in a region with a volume ΔV at time t. The state of i-th real-droplet is characterized by a set of variables $\boldsymbol{a}(t) = (a^{(1)}, a^{(2)}, \ldots, a^{(d)})$, where d is the number of independent variables. Hereafter, $\boldsymbol{a}_i(t)$ is referred to as the attribute of the i-th real-droplet. For example, following quantities can be an attribute: the equivalent radius, which represents the amount of water that the real-droplet contains, defined as the radius of a sphere having the same volume as the contained water; the mass of NaCl solute contained in the real-droplet. (In general, several types of soluble/insoluble aerosols are contained in each real-droplet); the electric charge of the real-droplet.

Two real-droplets may collide and coalesce into one bigger real-droplet. Let us denote the attribute of the newly created real-droplet through the coalescence of the real-droplets j and k as $\boldsymbol{a}' = \boldsymbol{a}_j + \boldsymbol{a}_k$. Note here that the new attribute $\boldsymbol{a}'$ cannot be expressed in a simple summation of $\boldsymbol{a}_j$ and $\boldsymbol{a}_k$ in general, but we employ this expression to simplify the notation. We consider that the real-droplets are always well-mixed inside the region with the volume

ΔV and the real-droplets coalesce with each other in a probabilistic manner. That is, there exists a probability that an arbitrary pair of real-droplets j and k will coalesce in a short time interval $(t, t + \Delta t)$, which is given by

$$P_{jk} = K(\boldsymbol{a}_j, \boldsymbol{a}_k)\frac{\Delta t}{\Delta V}, \tag{1}$$

here $K(\boldsymbol{a}_j, \boldsymbol{a}_k)$ is a certain function of $\boldsymbol{a}_j$ and $\boldsymbol{a}_k$.

There are various elementary processes in cloud microphysics, such as the sedimentation and advection of real-droplets, the condensation / evaporation of vapor, and the chemical reactions, but we do not consider these processes in this paper to make our discussion simple.

3 Super-Droplets

Let us introduce super-droplets to represent the stochastic coalescence of real-droplets.

A super-droplet is characterized by the attribute $\boldsymbol{a}$ and a multiplicity $\xi \in \{1, 2, \dots\}$. We consider that this super-droplet represents ξ number of real-droplets with the same attribute $\boldsymbol{a}$. Our idea is to approximate the real-droplets $\{\boldsymbol{a}_i | i = 1, 2, \dots, N_r\}$ by the super-droplets $\{(\boldsymbol{a}_i, \xi_i) | i = 1, 2, \dots, N_s\}$, $N_s < N_r$.

We define that the super-droplets j and k will coalesce in the following way.

1. If $\xi_j \neq \xi_k$, choosing j to satisfy $\xi_j > \xi_k$ without losing generality, and

$$\xi'_j = \xi_j - \xi_k, \quad \xi'_k = \xi_k, \tag{2}$$
$$\boldsymbol{a}'_j = \boldsymbol{a}_j, \quad \boldsymbol{a}'_k = \boldsymbol{a}_j + \boldsymbol{a}_k, \tag{3}$$

   where the dashed valuables represent the updated value after the coalescence.

2. If $\xi_j = \xi_k$,

$$\xi'_j = [\xi_j/2], \quad \xi'_k = \xi_j - [\xi_j/2], \tag{4}$$
$$\boldsymbol{a}'_j = \boldsymbol{a}'_k = \boldsymbol{a}_j + \boldsymbol{a}_k, \tag{5}$$

   where Gauss' symbol $[b]$ is the greatest integer that is less than or equal to the real number b.

Figure. 2 is a schematic view of the coalescence of super-droplets. In this example, $\xi_j = 3$ and $\xi_k = 2$. We can see that $\min(\xi_j, \xi_k) = 2$ pairs of real-droplets undergo coalescence, which results in the decrease of multiplicity $\xi_j : 3 \rightarrow 1$ and the increase of the size of the super-droplet k.

This definition of super-droplet coalescence possesses a favorable property that the number of super-droplets is unchanged in most cases though the number of real-droplets always decreases. The number of super-droplets decreases through the coalescence only when $\xi_j = \xi_k = 1$, i.e., both super-droplets are real-droplets. This results in $\xi'_j = 0$ and $\xi'_k = 1$, and we remove the super-droplet j out of the system. Hence, we can say that the number of super-droplets is conserved in most cases

$$N_s(t) \simeq \text{const.} \tag{6}$$

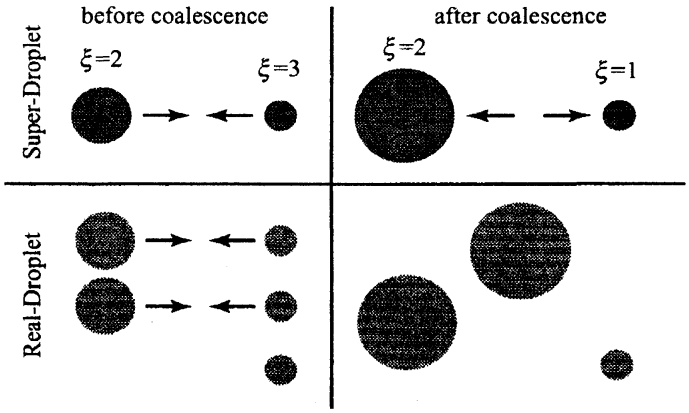

Figure 2: Schematic view of the coalescence of super-droplets. Two super-droplets with multiplicity 2 and 3 undergo coalescence (upper left). This represents the coalescence of two real-droplet pairs (lower left and right). As a result the super-droplet with multiplicity 2 becomes larger and the multiplicity of the other super-droplet decreases $3 \longrightarrow 1$ (upper right).

Because the number of super-droplets corresponds to the accuracy of SDM, the number conservation of super-droplets suggests the flexible response of SDM to the drastic change of the number of real-droplets.

We have defined how a pair of super-droplets coalesce. The probability that an arbitrary pair of super-droplets j and k will coalesce in a short time interval $(t, t + \Delta t)$ is given by

$$P_{jk}^{(s)} = \max(\xi_j, \xi_k) P_{jk}. \tag{7}$$

With this definition of coalescence probability, the expectation value becomes identical to that of the coalescence of real-droplets. The super-droplet j represents ξ_j number of real-droplets with attribute $\boldsymbol{a}_j$ and the super-droplet k represents ξ_k number of real-droplets with attribute $\boldsymbol{a}_k$. In terms of the real-droplet world, this corresponds to the situation that there are $\xi_j \xi_k$ number of real-droplet pairs which have the possibility to coalesce with the probability P_{jk}. Thus, the number of real-droplet pairs which will coalescence follows the binomial distribution with $\xi_j \xi_k$ trials and success probability P_{jk}. Thus, in the real-droplet world, the expectation value of the number of coalesced pairs is $\xi_j \xi_k P_{jk}$. On the other hand, in the super-droplet world, a coalescence of the super-droplets j and k represents the coalescence of $\min(\xi_j, \xi_k)$ pairs of real-droplets with attribute $\boldsymbol{a}_j$ and $\boldsymbol{a}_k$. Thus, the coalescence of the super-droplets j and k is expected to represent $\min(\xi_j, \xi_k) P_{jk}^{(s)} = \min(\xi_j, \xi_k) \max(\xi_j, \xi_k) P_{jk} = \xi_j \xi_k P_{jk}$ number of coalescence of real-droplet pairs, which is identical to the value in the real-droplet world. It is worth noticing here that the variance in the super-droplet world becomes larger than that in the real-droplet world. In the real-droplet world, the variance of the number of coalesced pairs is $\xi_j \xi_k P_{jk}(1 - P_{jk}) \simeq \xi_j \xi_k P_{jk} =: V_r$ (Poisson distribution limit). In the super-droplet world, the variance is $\{\min(\xi_j, \xi_k)\}^2 P_{jk}^{(s)} - \{\xi_j \xi_k P_{jk}\}^2 \simeq \min(\xi_j, \xi_k) V_r$, which is $\min(\xi_j, \xi_k)$ times larger than that of the real-droplet world.

The definition of the coalescence process of super-droplets is still theoretical and it can not be implemented on a computer directly. Based on the above definition, the present author and his coworkers have developed a novel Monte Carlo scheme to simulate the stochastic coalescence of super-droplets [16]. We do not show the precise calculation procedure in the present paper, but one thing worth emphasizing is that the *operation count* and *memory*

required for the computation of our Monte Carlo scheme is $O(N_s)$:

$$operation \sim N_s, \quad memory \sim N_s. \tag{8}$$

4 Computational cost of SDM

In this section we estimate the computational cost of SDM. Before discussing the computational cost, we have to clarify what is the quantity we want to reproduce or predict using SDM. Let us choose the number density distribution of real-droplets $n(\boldsymbol{a}, t)$ as the reference quantity and discuss how much computational cost is necessary to reproduce $n(\boldsymbol{a}, t)$ within a certain margin of error.

We have to estimate the number density distribution $n(\boldsymbol{a})$ from the super-droplets $\{(\xi_i, \boldsymbol{a}_i) \mid i = 1, 2, \ldots, N_s\}$. For the estimation, we use the kernel density estimation method, which was originally developed to estimate the generating probability distribution from its random sample [17].

We use the density estimator function $\tilde{n}(\boldsymbol{a})$ with Gaussian kernel $W_\sigma^{(d)}(\boldsymbol{a})$, defined by

$$\begin{aligned} \tilde{n}(\boldsymbol{a}) &:= \sum_{i=1}^{N_s} \xi_i W_\sigma^{(d)}(\boldsymbol{a} - \boldsymbol{a}_i), \\ W_\sigma^{(d)}(\boldsymbol{a}) &:= \frac{1}{(\sqrt{2\pi}\sigma)^d} \exp\left\{-\boldsymbol{a}^2/2\sigma^2\right\}. \end{aligned} \tag{9}$$

To evaluate the error let us introduce the Mean Integrated Squared Error (MISE) defined by

$$C(\sigma) = E\left[\int d^d a \ \{n(\boldsymbol{a}) - \tilde{n}(\boldsymbol{a})\}^2\right]. \tag{10}$$

Note here that $C(\sigma)$ is defined as an ensemble averaged value because each $\{(\xi_i, \boldsymbol{a}_i)\}$ is one of the random realizations of the stochastic coalescence process, and it measure the difference of $\tilde{n}(\boldsymbol{a})$ from the exact solution $n(\boldsymbol{a})$.

Before evaluating MISE (10), let us examine the behavior of the number density of super-droplets $q(\xi, \boldsymbol{a}, t; N_s)$, i.e., $q(\xi, \boldsymbol{a}, t; N_s)\Delta^d a$ is the expectation number of super-droplets with multiplicity ξ and attribute in the small interval $(\boldsymbol{a}, \boldsymbol{a} + \Delta^d a)$ at time t. Obviously, $q(\xi, \boldsymbol{a}, t; N_s)$ depends on the total number of super-droplets N_s. Let us assume that the following form of scaling law exists:

$$q(\xi, \boldsymbol{a}, t; \alpha N_s) = \alpha^{k_1} q(\alpha^{k_2}\xi, \boldsymbol{a}, t; N_s). \tag{11}$$

The equality below holds good by definition

$$\int d^d a \sum_{\xi=0}^{\infty} q(\xi, \boldsymbol{a}, t; N_s) = N_s \tag{12}$$

and remember that N_s is almost conservative in time. Also remember that super-droplets represent the expected dynamics of the real-droplets irrespective of the choice of N_s, which yields

$$\sum_{\xi=0}^{\infty} \xi q(\xi, \boldsymbol{a}, t; N_s) = n(\boldsymbol{a}, t). \tag{13}$$

Assuming that $q(\xi, \boldsymbol{a}, t; N_s)$ is smooth enough in ξ, the equality below may holds good,

$$\sum_{\xi=0,\alpha,2\alpha,\ldots}^{\infty} \simeq \frac{1}{\alpha} \sum_{\xi=0,1,2\ldots}^{\infty}, \tag{14}$$

for the summation of $q(\xi, \boldsymbol{a}, t; N_s)$. Then, from the equations (11)-(14) we can determine the scaling exponents as $(k_1, k_2) = (2, 1)$.

Based on this scaling law, we can derive the expectation value and variance of $\tilde{n}(\boldsymbol{a})$ as

$$E[\tilde{n}(\boldsymbol{a})] \simeq n(\boldsymbol{a}) + \frac{\sigma^2}{2}\left\{\sum_j \frac{\partial^2 n(\boldsymbol{a})}{\partial a_j^2}\right\},$$
$$V[\tilde{n}(\boldsymbol{a})] \simeq \frac{N_r n(\boldsymbol{a})}{N_s(2\sqrt{\pi}\sigma)^d}.$$

Substituting these two equations into (10), we can determine the σ^* that minimize the MISE $C(\sigma)$, which yields the scaling,

$$\sigma^* \sim N_s^{\frac{-1}{(d+4)}}, \quad C(\sigma^*) \sim N_s^{\frac{-4}{(d+4)}}.$$

Thus, the *operation count* and *memory* needed for SDM scales like

$$\begin{aligned} \textit{operation} \sim N_s &\sim \left(\frac{1}{\sqrt{C(\sigma^*)}}\right)^{(d+4)/2}, \\ \textit{memory} \sim N_s &\sim \left(\frac{1}{\sqrt{C(\sigma^*)}}\right)^{(d+4)/2}. \end{aligned} \tag{15}$$

We have estimated the computational cost of SDM. However, our result depends on our choice of the estimator kernel function $W_\sigma^{(d)}(\boldsymbol{a})$ though kernel density estimation method itself is irrelevant to SDM. In general, any function which meets some appropriate conditions can be chosen as an estimator kernel function, and Gaussian kernel, which we adopted, is a sort of order-2 kernel. Our estimation could be improved if we use higher order kernel, but do not discuss this point in detail in this paper.

5 Is SDM computationally more efficient than spectral (bin) method?

The computational cost of SDM has been estimated, and if the cost is lower than other cloud microphysics models, we can say that the use of SDM is beneficial. Let us compare the computational cost of SDM and spectral (bin) method in this section.

Spectral (bin) method is a sort of finite difference scheme to simulate the time evolution of $n(\boldsymbol{a}, t)$ [9–13]. From the governing law of the stochastic coalescence process (1) we can derive the time evolution equation of $n(\boldsymbol{a}, t)$:

$$\begin{aligned} \frac{\partial n}{\partial t} =& \frac{1}{2}\int d^d a' n(\boldsymbol{a}') n(\boldsymbol{a} - \boldsymbol{a}') K(\boldsymbol{a}', \boldsymbol{a} - \boldsymbol{a}') \\ & - n(\boldsymbol{a}) \int d^d a' n(\boldsymbol{a}') K(\boldsymbol{a}, \boldsymbol{a}'). \end{aligned} \tag{16}$$

This d-multiple integro-differential equation is called the Stochastic Coalescence Equation (SCE).

The basic idea of spectral (bin) method is to discretizing $n(\boldsymbol{a}, t)$ into a several number of bins (histograms) and simulate SCE. Let N_b be the number of bins for each attribute, i.e., the number of grid points used for the discretization of $n(\boldsymbol{a})$ per attribute. Then, because SCE is a d-multiple integro-differential equation, the *operation count* and *memory* required for the computation of spectral (bin) method scales like "*operation* $\sim N_b^{2d}$" and "*memory* $\sim N_b^d$."

The error evaluation function corresponding to MISE (10) is the Integrated Squared Error (ISE) defined by

$$C = \int d^d a \ \{n(\boldsymbol{a}) - n_b(\boldsymbol{a})\}^2 . \tag{17}$$

Here, $n_b(\boldsymbol{a})$ is the approximate solution generated by the spectral (bin) method. If the accuracy of spectral bin method is kth order in attribute-space, C scales like $C \sim N_b^{-2k}$ by definition.

Combining the above mentioned results, we can estimate the scaling of *operation count* and *memory* needed for the computation of spectral (bin) method in terms of C as follows,

$$\begin{aligned} \textit{operation} &\sim N_b^{2d} \sim \left(\frac{1}{\sqrt{C}}\right)^{2d/k}, \\ \textit{memory} &\sim N_b^{d} \sim \left(\frac{1}{\sqrt{C}}\right)^{d/k}. \end{aligned} \tag{18}$$

Now, we are ready to compare the computational efficiency of SDM and spectral (bin) method by comparing the exponents in (18) and (15). The result suggest that the operation count of SDM becomes lower than spectral (bin) method when the condition

$$d > \frac{4k}{4-k} \quad \text{and} \quad k < 4,$$

is satisfied, and the memory of SDM becomes lower than spectral (bin) method when

$$d > \frac{4k}{2-k} \quad \text{and} \quad k < 2.$$

Hence, if $k = 1$, i.e., the accuracy of the spectral (bin) method is 1st order, SDM is more efficient when the number of attributes $d \geq 2$. Our results also suggest that when $k \geq 4$, SDM is always not efficient. However, as already mentioned, our estimation depends on the choice of the estimator kernel function and the computational efficiency should be discussed more carefully. At least we could say that there is a tendency that SDM is computationally less demanding when the number of attributes d is large and the accuracy of the spectral (bin) method k is not so high-order.

6 Concluding remarks

SDM is a novel, particle-based, probabilistic microphysics model. In the present paper, we estimated the computational cost of SDM. To simplify the discussions, we focused our attention to simulate the stochastic coalescence process and neglected other elementary processes, such as sedimentation, evaporation/condensation. Based on a similar discussions

developed in the kernel density estimation theory [17], we estimated the relationship between the computational cost and computational error of SDM. Comparing the results with that of the spectral (bin) methods, it was suggested that SDM is computationally less demanding when the number of attributes d is large and the accuracy of the spectral (bin) method k is not so high-order.

In various research areas, many types of particle-based simulation schemes have been developed in these days. SDM can be regarded as a sort of Direct Simulation Monte Carlo (DSMC) method, which was initially proposed to simulate the Boltzmann equation for predicting rarefied gas flows [18]. Our discussions in this paper is based on (11)-(14), which are fundamental properties and assumptions of SDM and irrespective of the detail procedure of SDM. Consequently, a similar way of evaluation may be also applicable to other particle-based simulation schemes.

References

[1] M. Z. Jacobson, *Fundamentals of Atmospheric Modeling*, 2nd ed. (Cambridge Univ Press, New York, 2005), p. 813.

[2] E. Kessler, "On the Distribution and Continuity of Water Substance in Atmospheric Circulations," Meteor. Monogr **10,** 1–84 (1969).

[3] C. L. Ziegler, "Retrieval of thermal and microphysical variables in observed convective storms. Part I: Model development and preliminary testing," J. Atmos. Sci. **42,** 1487–1509 (1985).

[4] M. Murakami, "Numerical Modeling of Dynamical and Microphysical Evolution of an Isolated Convective Cloud," J. Meteor. Soc. Japan **68,** 107–128 (1990).

[5] B. S. Ferrier, "A double-moment multiple-phase four-class. bulk ice scheme. Part I: Description," J. Atmos. Sci. **51,** 249–280 (1994).

[6] M. P. Meyers, R. L. Walko, J. Y. Harrington, and W. R. Cotton, "New RAMS Cloud Microphysics Parameterization. Part II: The Two-Moment Scheme," Atmos. Res. **45,** 3–39 (1997).

[7] M. Khairoutdinov and Y. Kogan, "A New Cloud Physics Parameterization in a Large-Eddy Simulation Model of Marine Stratocumulus," Mon. Wea. Rev. **128,** 229–243 (2000).

[8] A. Seifert and K. D. Beheng, "A double-moment parameterization for simulating autoconversion, accretion and selfcollection," Atmos. Res. **59-60,** 265–281 (2001).

[9] E. X. Berry, "Cloud Droplet Growth by Collection," J. Atmos. Sci. **24,** 688–701 (1967).

[10] S.-T. Soong, "Numerical Simulation of Warm Rain Development in an Axisymmetric Cloud Model," J. Atmos. Sci. **31,** 1262–1285 (1974).

[11] A. Bott, "A Flux Method for the Numerical Solution of the Stochastic Collection Equation," J. Atmos. Sci. **55,** 2284–2293 (1998).

[12] A. Bott, "A Flux Method for the Numerical Solution of the Stochastic Collection Equation: Extension to Two-Dimensional Particle Distributions," J. Atmos. Sci. **57,** 284–294 (2000).

[13] A. Khain, A. Pokrovsky, M. Pinsky, A. Seifert, and V. Phillips, "Simulation of Effects of Atmospheric Aerosols on Deep Turbulent Convective Clouds Using a Spectral Microphysics Mixed-Phase Cumulus Cloud Model. Part I: Model Description and Possible Applications," J. Atmos. Sci. **61,** 2963–2982 (2004).

[14] D. T. Gillespie, "An Exact Method for Numerically Simulating the Stochastic Coalescence Process in a Cloud," J. Atoms. Sci. **32,** 1977–1989 (1975).

[15] M. Seeßelberg, T. Trautmann, and M. Thorn, "Stochastic simulations as a benchmark for mathematical methods solving the coalescence equation," Atmos. Res. **40,** 33–48 (1996).

[16] S. Shima, K. Kusano, A. Kawano, T. Sugiyama, and S. Kawahara, "Super-Droplet Method for the Numerical Simulation of Clouds and Precipitation: a Particle-Based Microphysics Model Coupled with Non-hydrostatic Model," arXiv:physics/0701103 .

[17] G. R. Terrell and D. W. Scott, "Variable Kernel Density Estimation," Ann. Statist. **20,** 1236–1265 (1992).

[18] G. A. Bird, *Molecular gas dynamics and the direct simulation of gas flows* (Clarendon Press, Oxford, 1994), p. 458.