Cloudy with a Chance of Microphysics: Modeling Droplet Collisions for the Climate Scale

Thesis by
Emily de Jong

In Partial Fulfillment of the Requirements for the
Degree of
Ph.D. in Mechanical Engineering

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2025
Defended June 18, 2024

ACKNOWLEDGEMENTS

My studies at Caltech were a dichotomy of working hard on research and courses, and playing hard in the mountains of Southern California and beyond. I am eternally grateful to those who supported and enabled both sides of my graduate school experience.

Most importantly, thank you to my parents for supporting my choice to go to graduate school in Lala-land. You have been accepting and supportive of my endeavors, both academic and personal, even as my interests and enthusiasm have ebbed and flowed and rapidly changed direction over the years.

Next, I thank the community of scientists, engineers, and fellow students in CliMA for being an unwavering source of support, both academically and personally. To my advisor Tapio, thank you for pushing me to be an independent thinker and researcher, all while being an unconditionally supportive mentor. To Clare, thank you for helping me navigate an unfamiliar field, for tolerating my unbridled and wacky sense of humor, and for being one of my greatest mentors and friends throughout graduate school and beyond. To Anna, Melanie, Toby, and Zhaoyi, thank you for guiding and contributing to my research. It has been a pleasure to share science, stories, and laughter around the coffee machine and over lunch in the garden throughout my time as a graduate student.

As a mechanical engineering student, I have been privileged to have an additional academic community beyond my research. To Tracy, Alexandra, Helen, and the other women in MCE, thank you for always being there to commiserate, comfort, and celebrate each other. To Ruby, thank you for being a mentor and a role model among the faculty. To Guillaume, thank you for guiding my first year at Caltech and encouraging me to explore early on. This community extends far beyond Caltech as well. To the Krell Institute and the CSGF network, thank you for being a resource for personal and professional development. To Sylwester, thank you for teaching me how to be a better scientific software developer, for promoting my work and treating me as a peer, and for working so hard to advance your field and those around you. To Eliot, thank you for being a present and involved mentor at NREL, for giving me the guidance to succeed on an unfamiliar topic but also the freedom to explore. To Hugh, Marcus, Adele, and Kara, thank you for welcoming me to the broader microphysics community.

Finally, thank you to those who have kept me safe, sane, and happy outside of the academic realm. This includes many friends and adventure partners from the Caltech Alpine Club, in addition to cycling/skiing/climbing/running buddies further away. Thank you Marty for being a supportive partner in more ways than one: few people are able to climb a mountain in the morning, help edit my essays in the evening, and still party all night. Thank you to my friends Ruchita, Rushy, George, Shayal, and Stratos for cheering me on from afar and for being only a phone call away. Thank you to Hannah, Sophia, Britnee, Livia, and Elle for having fun with me on bicycles, tennis courts, trails, and dance floors. While there are too many names to list in entirety, I will remember fondly many times spent with friends on ski trips, climbing trips, bike rides, beach excursions, and other mountain shenanigans. I feel fortunate to have found such an impressive, kind, and talented set of individuals to spend time with outside of my studies.

ABSTRACT

Feedbacks between a warming atmosphere, emission of aerosols, and clouds and precipitation are some of the most difficult aspects for climate models to accurately capture. While climate models operate at resolutions of tens or hundreds of kilometers, many of the physics that determine how and where clouds form or precipitate function at the micron droplet scale. Due to this disparity in physical scales, most of these cloud physics must be modeled with only a few approximate quantities and physical equations. These simplifications lead to large uncertainties about climate forcings such as the sensitivity of global warming to human-emitted aerosols.

This work presents several promising new techniques for modeling and understanding hydrometeors in the climate system, with a particular focus on processes that involve collisions between droplets. First, I extend a high-complexity high-fidelity Lagrangian microphysics method to represent the process of breakup, in which colliding droplets fragment upon collision. Next, I introduce two new methods which attempt to reduce the assumptions inherent to modeling droplet coalescence, in which colliding droplets combine to form a larger drop. The first method uses a spectral finite element approach, while the second generalizes this technique using a method of moments to create a fully flexible microphysics scheme. Finally, I turn to remote observations of clouds, aerosols, and lightning over busy shipping regions to offer new techniques for quantifying aerosol-cloud interactions from creative data resources. This combination of high-fidelity modeling tools, observational data, and efficient numerical methods offers a path toward improving our understanding of the role of cloud microphysics in our climate system.

PUBLISHED CONTENT AND CONTRIBUTIONS

de Jong, E. K., Schneider, T., Jalkanen, J.-P., & Wennberg, P. O. (2024). What can lightning and shipping regulations tell us about aerosols in deeply convecting clouds? *ESS Open Archive; Submitted to Geophysical Research Letters*. https://doi.org/10.22541/essoar.171632547.79526166/v1

E.K.dJ participated in the conception of the project and led the experimental design, analysis, and writing.

de Jong, E. K., Mackay, J. B., Bulenok, O., Jaruga, A., & Arabas, S. (2023). Breakups are complicated: Breakups are complicated: An efficient representation of collisional breakup in the superdroplet method. *Geoscientific Model Development*, *16*(14), 4193–4211. https://doi.org/10.5194/gmd-16-4193-2023

E.K.dJ led the design of project, implementation of the algorithm, and writing of the manuscript, and participated in the analysis and demonstration of the algorithm's accuracy and convergence.

de Jong, E. K., Bischoff, T., Nadim, A., & Schneider, T. (2022). Spanning the gap from bulk to bin: A novel spectral microphysics method. *Journal of Advances in Modeling Earth Systems*, *14*(11), e2022MS003186. https://doi.org/10.1029/2022MS003186

E.K.dJ participated in the conception of the project and development of the core mathematical methods, and led the computational implementation, demonstration and analysis, and writing of the manuscript.

TABLE OF CONTENTS

FOREWORD

The following essay was my (unsuccessful) submission to the 2024 DOE CSGF "Communicate Your Science and Engineering" competition. Beyond its entertainment value, I hope it may provide context for my journey pursuing a Mechanical Engineering PhD by studying cloud microphysics.

Two cloud droplets walk into a bar; do they make it rain?

Five years ago, I never imagined making such a niche science pun. Having recently finished my second internship in the oil industry as a chemical engineer, I began my PhD at Caltech, excited to study soot and combustion products.

I never imagined that I would one day call myself a climate scientist, yet here we are.

As it turns out, soot particles and cloud droplets share a lot in common. Both float around in the atmosphere and are related to human impacts on climate. Both provide endless opportunities for scientific play-on-words (“Your simulation must be huge, the HPC cluster is smoking”). And, interestingly enough, soot and raindrops both form by the same physical process of coagulation, also known as “collisional coalescence”.

Let’s unpack the introductory pun to help explain this common physical mechanism. Clouds are collections of tiny water droplets suspended in the air, droplets which grow and shrink by condensing and evaporating water from the surrounding air. When these droplets become large enough, they collide with each other (i.e. walk into the same metaphorical bar). When these collisions happen, droplets often stick together and coalesce, forming an even larger droplet. Soot forms via the same process: small aromatic hydrocarbons produced from combustion bump into each other and coagulate to form a large soot aggregate. Soot aggregates cause lung cancer, clog up engines, and contribute to global warming. Coalescing cloud droplets make it rain.

As a first-year mechanical engineering PhD student, I knew none of these fun facts about the micro-scale physics of clouds. Fresh out of coursework on hypersonic fluid mechanics and exploring potential research groups, I was intrigued by an opportunity to branch into atmospheric fluid mechanics. I soon learned that my would-be advisor Professor Tapio Schneider had an ambitious goal in mind: to revolutionize the way we model climate.

When it came time to choose between Tapio's research and my initial plan to study combustion products, I knew next to nothing about the basics of Earth's atmosphere. Even worse, Covid-19 upended the world, terminating the in-person meetings with postdocs and other graduate students that were my only lifeline to making sense of this new scientific field. When Caltech shut down the campus, I drove from Pasadena to Colorado to weather the storm with family. Later that same week, I received the call offering me a graduate fellowship (the CSGF). Instilled with the urgency of worldwide crisis and presented with this opportunity, my ultimate decision was easy. This was my chance to help solve a different global crisis: the uncertainty of climate change.

In the ensuing year, I learned about Earth's atmosphere largely by osmosis, attending journal clubs on climate dynamics ("to Rossby, or not to Ross-be?") and drinking from the firehose of online classes on Atmospheric Chemistry ("Matlab's ODE23-and-me") and Clouds and Boundary Layer Dynamics ("so flam-buoyant"). At the same time, I dove into my own initial goal of building an aerosol model for the CliMA project. I brushed up on mathematical techniques like statistical distributions and the "method of moments," all while continuously discovering flaws in the way that we model not only aerosol particles in the atmosphere, but also every cloud droplet, ice crystal, and snowflake. If our goal was to reduce uncertainty in this new climate model, I decided that we would need to rebuild our computational representation of these cloud microphysics from the ground up.

Modeling the atmosphere is conceptually similar to other applications of computational fluid mechanics. Large-scale motions like average winds are fully resolved by dividing the atmosphere into gridboxes, while smaller motions at the scale of turbulence and clouds are parameterized at the "sub-grid" scale. My metaphorical watering-hole of cloud and rain drops exists at infinitesimal scales compared with these 10-100km grid boxes, yet these microphysics determine important climatological details such as where and how much it rains, or how human-emitted aerosols impact climate. To produce accurate and confident climate predictions, we must design the underlying mathematical systems such that our supercomputers are efficient at not only moving information from gridbox to gridbox, but also at solving complex systems of equations and thermodynamics at the subgrid scale.

We typically represent cloud particles as "distributions": rather than tracking individual droplets directly, we consider the probability of finding a drop of a particular size within the gridbox. The equations governing how these droplet populations evolve due to condensation, coalescence, and other processes depend on droplet size, adding a fifth dimension to our already four-dimensional space-time atmospheric system. My first PhD paper attacked the numerical challenges and efficiency of this issue head-on, presenting methods to solve for a droplet-size distribution without making the sweeping and inaccurate assumptions of canonical models.

My fascination with droplet collisions drew me to more detailed models of cloud microphysics based on Monte Carlo methods. These models still require heavy computational power but operate at the scale of tens or hundreds or meters to model a single cloud. As a project for my class in GPU computing, I decided to answer an outstanding question related to a different collisional process: what happens when droplets collide, and then break-up rather than coalescing? The answer is that this break-up process has little impact on cloud properties compared to coalescence, but it put to rest an often-cited source of uncertainty in cloud microphysics and climate. More recently, I turned my attention to satellite observations to inform our models of cloud microphysics. For instance, collisions between ice crystals are poorly understood compared with their liquid droplet cousins, but these collisions produce interesting evidence in the form of charge transfer and ultimately lightning.

As a climate scientist, I found my niche in this peculiar corner of atmospheric modeling, where collisions between micron-sized particles can have flashy and significant impacts on global climate. And as an engineer, I love approaching these modeling challenges by drawing on my problem-solving skills to think outside of the gridbox.

Chapter 1

INTRODUCTION

1.1 Clouds and Climate

Clouds provide a crucial link between human action and climate reaction. As humans pump greenhouse gases and other pollutants into the atmosphere, our climate responds in predictable ways, i.e. warming according to the greenhouse gas effect. But the atmosphere also responds in less predictable ways thanks to the complexity of clouds. Emissions from ships traversing the seas form streaks of brighter clouds, which are visible from space and which have a global cooling effect (Christensen et al., 2022). Convective clouds above wildfires and heavily polluted industrial regions can be strengthened by these aerosols, becoming electrified storm clouds (Liu et al., 2021); other times, these aerosols have the opposite impact, delaying or preventing these systems from precipitating (Igel & van den Heever, 2021; Varble et al., 2023). Present-day climate models cannot accurately or consistently predict these changes, particularly how the clouds which cool our planet respond to changes in human-emitted aerosols, and especially not on the timescales necessary for policy decisions about climate change (Intergovernmental Panel on Climate Change, 2014; Morrison et al., 2020).

At the root of this challenge is a separation of scales. Climate models simulate timescales of years and resolve atmospheric dynamics at the scale of 10s or 100s of kilometers. On the other hand, the physics that determine when, where, and how clouds form, evolve, and precipitate operate at timescales of seconds and length scales of microns (10^{-6}m). Consider a stratocumulus cloud looming above Caltech's (approximately) 1-square-kilometer campus on a gloomy June day: this thin low cloud might have a depth of 200m and a density of 100 cloud droplets per cm^3. That means a total of 2×10^{16} cloud droplets exist in the cloud above Caltech's campus, all of which are constantly condensing and evaporating water, colliding with each other, and falling relative to the air around them. Scaled up to a region the size of Los Angeles, California, or the USA, even the world's largest supercomputers cannot simulate this many cloud particles (or "hydrometeors"), much less at the scale of global climate.

1.1.1 Challenges in cloud microphysics

As with other underresolved atmospheric processes such as turbulence, the separation of scales between cloud microphysics and model resolution is addressed through a set of "subgrid scale" parameterizations. Rather than directly simulating the formation, growth, and evolution of individual cloud particles, we treat all hydrometeors within each 10-100km atmospheric gridboxes as belonging to a “population.” Most commonly, this population is described and tracked according to a few bulk properties such as the total mass of water belonging to these hydrometeors or the total number of particles (e.g. Kessler, 1969; Seifert and Beheng, 2006). These simplifying assumptions introduce structural uncertainty to the model. Furthermore, the process rates governing how hydrometeors evolve must be adapted to the simplified variables used to describe the population, leading to additional parametric uncertainty. More detailed methods exist, such as representing and evolving several characteristic members of the population through a technique called the superdroplet method (Shima et al., 2009) (described in Chapter 2), yet they are too computationally intensive for weather and climate predictions. Chapters 3 and 4 present alternative model structures that require similar or only slightly more computational resources than standard methods, but with the advantage of requiring fewer simplifying assumptions and parameters.

The other fundamental challenge to accurately representing cloud microphysics is a lack of adequate observational data to train and evaluate these subgrid scale models. Observations of hydrometeors at the individual particle scale are the most informative for fixing the parameters of microphysics process rates, yet these laboratory-scale measurements are challenging to perform and not necessarily representative of the complex atmospheric state. On the other hand, measurements of the atmospheric state can be readily attained from ground-based (radar or lidar) and space-based (satellite) measurements, but not at the spatiotemporal resolution required to distinguish microphysics from other subgrid-scale atmospheric processes. Nevertheless, these remote measurements offer the most complete history of clouds in the climate system and are crucial for evaluating larger scale climate sensitivity (e.g. Bellouin et al., 2020). In Chapter 5, I shift to focus on exactly this: creative ways to leverage available observational data in order to constrain aerosol-cloud interactions.

1.1.2 Collisional processes in clouds

One set of microphysical processes is particularly challenging to both model and measure: collisions between hydrometeors. When two liquid droplets collide with each other, they might merge ("coalesce") to form a larger droplet, they might fragment ("breakup") into many smaller droplets, or they could bounce elastically. In mixed phase clouds which include ice particles, the list of potential outcomes grows dramatically, with the splintering of ice particles upon collision believed to an important process in deep clouds (Phillips et al., 2017). This dissertation focuses on the first set of "warm-rain" collisional processes in Chapters 2-4 and touches on mixed-phase collisions in Chapter 5. Warm-rain collisions alone are a primary mechanism by which cloud droplets become large enough to precipitate: small droplets coalesce with their neighbors, forming larger and larger drops which collide faster and faster, eventually leading to rain. Parameterizations of these collisions in most climate models are lacking, leading to known issues such as an overprediction of drizzle from stratocumulus (Stephens et al., 2010).

While single-particle processes such as condensation, evaporation, and freezing of hydrometeors are well-constrained by thermodynamics, the rates and outcomes of collisional processes depend on exactly when, where, and how hydrometeors are moving toward or away from each other, as well as their properties. For instance, a large rain-sized droplet might fall faster relative to a smaller cloud-sized droplet beneath it, ultimately colliding and collecting the cloud droplet as it falls. If this size difference is large enough, the opposite could happen instead: the surface tension of the smaller droplet could be enough to prevent it from merging with the larger droplet, leading to a breakup process that leaves even more tiny droplet fragments in the wake of the collision (Low & List, 1982; Straub et al., 2010). When hydrometeors are viewed as a population with a distribution of radii or other properties, these collisional dynamics are best described as a set of probabilities: the probability that droplet A and droplet B collide given their relative sizes, or the probability that A and B coalesce versus breaking up. The coarsest microphysics models retain the least detail about the distribution of sizes or other population properties; thus they lead to the most assumptions, errors, and uncertainties regarding these crucial rain-forming processes.

1.2 Outline

Motivated by practical solutions for climate modeling, this dissertation presents new techniques for representing and understanding these small scale cloud processes (“cloud microphysics”), emphasizing collisional processes. The chapters progress from the smallest scale and most detailed simulations of microphysical processes, to simpler methods suitable for climate models, and finally a large-scale assessment of the sensitivity of clouds to aerosols.

Chapter 2 investigates the process of collisional breakup using a high-fidelity modeling technique known as the “superdroplet method.” This method tracks several representative droplets (a subset of the population) and treats collisional coalescences as random events with stochastic Monte Carlo steps (Shima et al., 2009). We introduce a complementary and efficient algorithm to represent collisional breakup, a process that was historically missing from this modeling framework. The chapter goes on to illustrate the impacts of collisional breakup on precipitation and cloud formation in a set of idealized cloud simulations. While the superdroplet method is too computationally expensive to run at a global scale, this work illustrates how it can be used to gain a process-level understanding that can inform cheaper climate-oriented methods.

The next two chapters return to simpler computational methods that are suitable for climate modeling, presenting new approaches that attempt to relieve limiting assumptions of standard bulk microphysics methods. Chapter 3 presents a novel numerical approach that involves directly modeling a distribution of particle sizes using collocated finite elements. This method can attain the accuracy of a more computationally expensive spectral method with only a quarter of the cost. The method presented in Chapter 4 builds on this distribution modeling approach by marrying the Eulerian method-of-moments with a flexible Lagrangian approach. Importantly, this method relaxes the assumption of most current models that all liquid hydrometeors are classified as either “cloud” or “rain” by applying a superdroplet-style approach to collisional processes between droplets in different subpopulations. As a result, we generalize the typical two-category moment-based method to utilize an arbitrary number of subdistributions while also eliminating the need for uncertain parameterizations of conversion rates between these categories.

Chapter 5 takes a step back from bottom-up mathematical modeling of microphysical processes to instead investigate large-scale perturbations to clouds in the climate system using remote observations. When new regulations in 2020 changed the fuels burned by overseas ship traffic, they also resulted in changes to the aerosols emitted by these ships, and therefore changes to the clouds above (Diamond, 2023; Yuan et al., 2022). This chapter investigates a new potential indicator of aerosol-cloud interactions: lightning, which comes from charge separation during ice particle collisions. This unusual metric reflects changes to shipping aerosols in regions which are typically overlooked by other aerosol-cloud assessments, providing a novel way to measure the impact of policy decisions on clouds and our climate. Finally, Chapter 6 offers concluding remarks to contextualize the scope and findings of this dissertation as well as suggest avenues for future work and improvement.

References

Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., Stier, P., Watson-Parris, D., Boucher, O., Carslaw, K. S., Christensen, M., Daniau, A.-L., Dufresne, J.-L., Feingold, G., Fiedler, S., Forster, P., Gettelman, A., Haywood, J. M., Lohmann, U., Malavelle, F., Mauritsen, T., . . . Stevens, B. (2020). Bounding Global Aerosol Radiative Forcing of Climate Change. *Reviews of Geophysics*, *58*(1), e2019RG000660. https://doi.org/10.1029/2019RG000660

Christensen, M. W., Gettelman, A., Cermak, J., Dagan, G., Diamond, M., Douglas, A., Feingold, G., Glassmeier, F., Goren, T., Grosvenor, D. P., Gryspeerdt, E., Kahn, R., Li, Z., Ma, P.-L., Malavelle, F., McCoy, I. L., McCoy, D. T., McFarquhar, G., Mülmenstädt, J., . . . Yuan, T. (2022). Opportunistic experiments to constrain aerosol effective radiative forcing [Publisher: Copernicus GmbH]. *Atmospheric Chemistry and Physics*, *22*(1), 641–674. https://doi.org/10.5194/acp-22-641-2022

Diamond, M. S. (2023). Detection of large-scale cloud microphysical changes within a major shipping corridor after implementation of the International Maritime Organization 2020 fuel sulfur regulations. *Atmospheric Chemistry and Physics*, *23*(14), 8259–8269. https://doi.org/10.5194/acp-23-8259-2023

Igel, A. L., & van den Heever, S. C. (2021). Invigoration or Enervation of Convective Clouds by Aerosols? *Geophysical Research Letters*, *48*(16), e2021GL093804. https://doi.org/10.1029/2021GL093804

Intergovernmental Panel on Climate Change. (2014). *Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [doi: 10.1017/CBO9781107415324]. Cambridge University Press.

Kessler, E. (1969). On the Distribution and Continuity of Water Substance in Atmospheric Circulations [doi: 10.1007/978-1-935704-36-2_1]. In E. Kessler (Ed.), *On the Distribution and Continuity of Water Substance in Atmospheric Circulations* (pp. 1–84). American Meteorological Society.

Liu, Y., Williams, E., Li, Z., Guha, A., Lapierre, J., Stock, M., Heckman, S., Zhang, Y., & DiGangi, E. (2021). Lightning Enhancement in Moist Convection With Smoke-Laden Air Advected From Australian Wildfires. *Geophysi-*

cal Research Letters, *48*(11), e2020GL092355. https://doi.org/10.1029/2020GL092355

Low, T. B., & List, R. (1982). Collision, Coalescence and Breakup of Raindrops. Part I: Experimentally Established Coalescence Efficiencies and Fragment Size Distributions in Breakup. *Journal of the Atmospheric Sciences*, *39*(7), 1591–1606. https://doi.org/10.1175/1520-0469(1982)039<1591:CCABOR>2.0.CO;2

Morrison, H., van Lier-Walqui, M. v., Fridlind, A. M., Grabowski, W. W., Harrington, J. Y., Hoose, C., Korolev, A., Kumjian, M. R., Milbrandt, J. A., Pawlowska, H., Posselt, D. J., Prat, O. P., Reimel, K. J., Shima, S.-I., Diedenhoven, B. v., & Xue, L. (2020). Confronting the Challenge of Modeling Cloud and Precipitation Microphysics. *Journal of Advances in Modeling Earth Systems*, *12*(8). https://doi.org/10.1029/2019MS001689
e2019MS001689 2019MS001689.

Phillips, V. T. J., Yano, J.-I., & Khain, A. (2017). Ice Multiplication by Breakup in Ice–Ice Collisions. Part I: Theoretical Formulation. *Journal of the Atmospheric Sciences*, *74*(6), 1705–1719. https://doi.org/10.1175/JAS-D-16-0224.1

Seifert, A., & Beheng, K. D. (2006). A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 1: Model description. *Meteorology and Atmospheric Physics*, *92*(1-2), 45–66.

Shima, S.-I., Kusano, K., Kawano, A., Sugiyama, T., & Kawahara, S. (2009). The super-droplet method for the numerical simulation of clouds and precipitation: A particle-based and probabilistic microphysics model coupled with a non-hydrostatic model. *Quarterly Journal of the Royal Meteorological Society*, *135*(642), 1307–1320. https://doi.org/10.1002/qj.441

Stephens, G. L., L'Ecuyer, T., Forbes, R., Gettelmen, A., Golaz, J.-C., Bodas-Salcedo, A., Suzuki, K., Gabriel, P., & Haynes, J. (2010). Dreary state of precipitation in global models. *Journal of Geophysical Research: Atmospheres*, *115*(24).

Straub, W., Beheng, K. D., Seifert, A., Schlottke, J., & Weigand, B. (2010). Numerical Investigation of Collision-Induced Breakup of Raindrops. Part II: Parameterizations of Coalescence Efficiencies and Fragment Size Distributions. *Journal of the Atmospheric Sciences*, *67*(3), 576–588. https://doi.org/10.1175/2009JAS3175.1

Varble, A. C., Igel, A. L., Morrison, H., Grabowski, W. W., & Lebo, Z. J. (2023). Opinion: A Critical Evaluation of the Evidence for Aerosol Invigoration of Deep Convection. *EGUsphere*, 1–31. https://doi.org/10.5194/egusphere-2023-938

Yuan, T., Song, H., Wood, R., Wang, C., Oreopoulos, L., Platnick, S. E., von Hippel, S., Meyer, K., Light, S., & Wilcox, E. (2022). Global reduction in ship-tracks from sulfur regulations for shipping fuel. *Science Advances*, *8*(29), eabn7988. https://doi.org/10.1126/sciadv.abn7988

Chapter 2

AN EFFICIENT REPRESENTATION OF COLLISIONAL BREAKUP IN THE SUPERDROPLET METHOD

de Jong, E. K., Mackay, J. B., Bulenok, O., Jaruga, A., & Arabas, S. (2023). Breakups are complicated: Breakups are complicated: An efficient representation of collisional breakup in the superdroplet method. *Geoscientific Model Development*, *16*(14), 4193–4211. https://doi.org/10.5194/gmd-16-4193-2023

Abstract

A key constraint of particle-based methods for modeling cloud microphysics is the conservation of total particle number, which is required for computational tractability. The process of collisional breakup poses a particular challenge to this framework, as breakup events often produce many droplet fragments of varying sizes, which would require creating new particles in the system. This work introduces a representation of collisional breakup in the so-called "superdroplet" method which conserves the total number of superdroplets in the system. This representation extends an existing stochastic collisional-coalescence scheme and samples from a fragment-size distribution in an additional Monte Carlo step. This method is demonstrated in a set of idealized box model and single-column warm-rain simulations. We further discuss the effects of the breakup dynamic and fragment-size distribution on the particle size distribution, hydrometeor population, and microphysical process rates. Box model experiments serve to characterize the impacts of properties such as coalescence effieciency and fragmentation function on the relative roles of collisional breakup and coalescence. The results demonstrate that this representation of collisional breakup can produce a stationary particle-size distribution, in which breakup and coalescence rates are approximately equal, and that it recovers expected behavior such as a reduction in precipitate-sized particles in the column model. The breakup algorithm presented here contributes to an open-source pythonic implementation of the superdroplet method, 'PySDM', which will facilitate future research using particle-based microphysics.

2.1 Introduction

The superdroplet method (SDM) for cloud microphysics is a high-fidelity particle-based (Lagrangian) representation of aerosols and hydrometeors that offers notable advantages over traditional bulk and bin microphysics schemes. Particle-based methods were initially used in atmospheric simulations to represent ice nucleation (Jensen & Pfister, 2004; Paoli et al., 2004; Shirgaonkar & Lele, 2006; Sölch & Kärcher, 2010), and were later extended to study aerosol indirect effects with a superdroplet approach (Andrejczuk et al., 2008) in which each "superdroplet" represents a mutliplicity of modeled particles with identical attributes, such as size and chemical properties. Later, the SDM was extended to include a stochastic representation of collisional coalescence (Riechelmann et al., 2012; Shima et al., 2009) and ice-phase processes (Shima et al., 2020), making the SDM a nearly-complete Monte Carlo representation of cloud microphysics. The burgeoning field of particle-based cloud microphysics uses SDM implementations in large-eddy simulations (LES) to understand microphysical processes that are underresolved in traditional bulk and bin methods (e.g., Andrejczuk et al., 2010; Chandrakar et al., 2021; Dziekan et al., 2019; Grabowski, 2020; Hoffmann, 2017; Morrison et al., 2019). Furthermore, a growing literature of machine learning in microphysics utilizes the SDM as a source of high-fidelity training data from which to "learn" microphysical tendencies and properties (Bieli et al., 2022; Seifert & Rasp, 2020). However, without a complete representation of microphysical processes in the SDM, its predictive and benchmarking power for cloud feedbacks is limited.

Many implementations of the SDM do not include the process of collisional breakup of droplets. Not only is collisional breakup an uncertain process in existing bin and bulk parameterizations (Grabowski et al., 2019; Morrison et al., 2020), but these uncertainties have been found to impact rain rates and other macroscale quantities in bin microphysics studies (Seifert et al., 2005). Studying collisional breakup in the SDM is not straightforward, as it requires balancing computational complexity within a mass-conserving numerical model that respects the physics of the process. Notably, a single breakup event is likely to produce fragments of multiple different sizes. A literal representation of all fragments in the SDM would require the creation of new superdroplet tracers for each new droplet size resulting from breakup, which can lead to an explosive growth of superdroplet quantity and dramatically inhibit performance of the SDM. Another option that reduces the computational burden of new superdroplets involves creating and then selectively merging superdroplets, as in Bringi et al., 2020. While both options respect mass conservation and the physics of superdroplet breakup, a scalable adaption of the SDM for parallel applications such as LES requires strict conservation of the total number of superdroplets. This work proposes a superdroplet-conserving SDM algorithm for the process of collisional breakup, conceptually similar to the mass-flux algorithm of Kotalczyk et al., 2017, using a Monte Carlo step that samples from a fragment size distribution.

This superdroplet-conserving breakup implementation draws inspiration from an analogous "superparticle" representation of phytoplankton (Jokulsdottir & Archer, 2016): individual phytoplankton aggregates spontaneously break uniformly into a number of fragments determined by a power law probability distribution. We apply a similar spontaneous breakup principle to an intermediate coalesced state resulting from the collision of two droplets. (While spontaneous breakup of liquid water droplets has also been investigated (Kamra et al., 1991), it has not been observed in in-situ studies of droplet collisions (Testik & Rahman, 2017) and is not included in this version of the SDM.) The presented collisional breakup algorithm utilizes empirical collection/breakup efficiencies (such as Beard and Ochs (1995), Berry (1967), and Schlottke et al. (2010)) to determine whether a colliding droplet pair is likely to break-up, and then samples from a corresponding empirical fragment size distribution (such as Beard and Ochs (1995), Low and List (1982), McFarquhar (2004), and Schlottke et al. (2010)) to determine the properties of the resulting fragmented superdroplet. Breakup parameterizations are typically complex and aim to summarize multiple physical mechanisms of breakup. This work addresses how the proposed SDM breakup algorithm samples from such complex fragment size distributions, but leaves evaluation and analysis of these empirical distributions to future work.

The contents of this paper proceed as follows: Section 2.2 begin with a conceptual description of the proposed breakup algorithm, followed by a mathematical description of its implementation. Section 2.3 validates the implementation against analytical results in the literature and explores convergence with number of superdroplets. Section 2.4 then presents several idealized simulations including various parameterizations of collisional breakup to demonstrate the behavior of this implementation in the SDM. Section 2.5 concludes the discussion and poses additional scientific questions which may be within reach given this novel implementation.

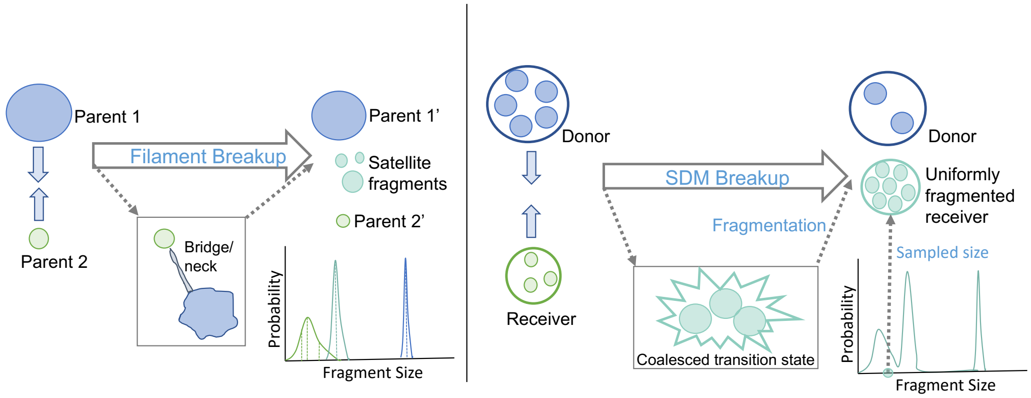


Figure 2.1: Conceptual view of a real filament breakup event (left) and the superdroplet-number-conserving SDM collisional breakup dynamic described in this work (right). The real event involves collision between two parent droplets, which may form a neck bridging each other before fragmenting into several differently sized droplets. The resulting droplets consisting of larger fragments derived from the initial colliding parents, plus a set of small fragments known as "satellites." Each of these groups forms a subdistribution in the overall fragmentation function, shown below the droplet schematics. The tracer-conserving SDM representation involves collision of two groups of droplets (each group represented as one superdroplet, a donor and receiver), which collide and coalescence into a transition state, which then fragments uniformly to a size sampled from the same fragmentation function as in the real case. The result of the SDM breakup is two superdroplets, or two groups of droplets, with one group corresponding to leftover donor droplets, and the other group corresponding to a set of fragments whose size may correspond to the depleted parent droplet size or the satellite fragments, depending on the fragment size sampling step.

2.2 Superdroplet-conserving Collisional Breakup

2.2.1 Conceptual description

Two colliding liquid hydrometeors in the atmosphere can break-up via several physical pathways, including filament, sheet, and disc breakup (Barros et al., 2008). The colliding droplets, referred to as "parents", typically lose mass to newly-formed tiny "satellite" droplets that result from the collision, thereby resulting in several differently sized droplet fragments (see Figure 2.1, left). As noted previously, computational scaling of the SDM relies on preserving the number of tracers in the system. In order to preserve the number of superdroplets in a binary collisional breakup event, breakup is treated as a two step process based on superdroplet-conserving coalescence (Figure 2.1, right). First, the two superdroplets collide and coalesce: the superdroplet of higher multiplicity acts as a "donor" by donating mass

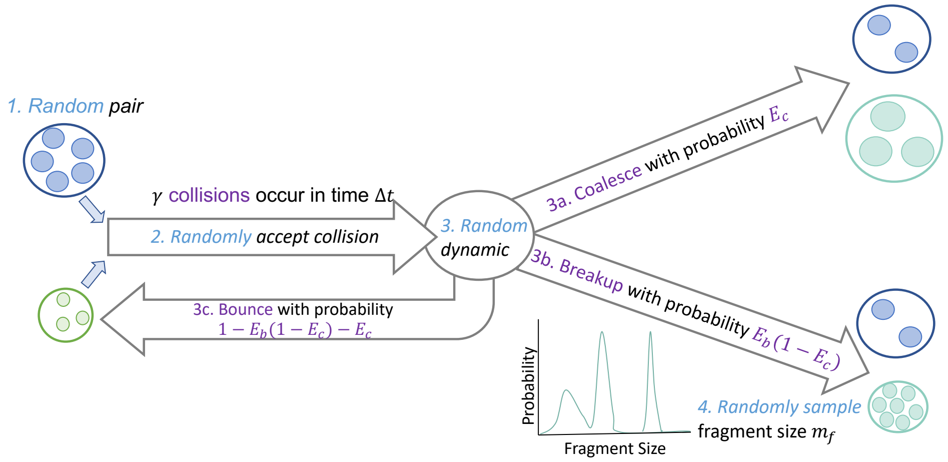


Figure 2.2: Diagram of the Monte Carlo decision pathway during a collision-coalescence-breakup event in the proposed algorithm.

and multiplicity while maintaining its attributes; the other superdroplet acts as a "receiver" by growing in mass and maintaining its multiplicity to form a "coalesced transition state." This unstable coalesced transition state immediately breaks up into fragments of uniform size: the fragment size is selected sampling from a distribution of fragment sizes that encompasses both the remnants of the original parent droplets, as well as the distribution of satellite fragments that can result from the collision. (Going forward, the term "fragment" will be used to describe all collisional breakup products, both small satellite fragments as well as larger fragments that are nearer in size to the original colliding droplets.) The attributes and multiplicity of this fragmented receiver are updated, with multiplicity increasing and mass of the individual droplets represented by the superdroplet decreasing. Uniform fragmentation is required to maintain conservation of superdroplets. Furthermore, uniform fragmentation requires the assumption that all superparticle attributes are extensive quantities and undergo equipartitional splitting (not applicable, e.g., for insoluble aerosol constituents). The product of a collisional breakup event is therefore two superdroplets: the donor maintains its attributes but donates multiplicity, and the fragmented receiver represents (uniform) fragments that result from the breakup event following a coalesced transition state. As in the original Monte Carlo step that determines whether a collision occurs, the fragment size is sampled at random from a fragment size distribution, which may depend on the properties of the colliding particles.

2.2.2 Mathematical description

The superdroplet-conserving method of collisional breakup is illustrated in Figure 2.2 and formulated below using notation following work of Shima et al. A single superdroplet with label i has a position $\mathbf{x}_i(t)$ and extensive physical attributes $\mathbf{a}_i(t)$, such as droplet volume or mass $(v_i(t), m_i(t))$ or mass of solute $(M_i(t))$. (Note that in Shima et al.'s notation, $\mathbf{a}_i(t)$ includes attributes such as droplet radius, whereas we only consider attributes which are linearly additive and extensive in the droplet size, such as volume or mass.) For simplicity, we will generally group all such extensive attributes together as $\mathbf{a}_i(t)$, but will specifically use the droplet mass m_i in computations of the transfer of extensive properties between superdroplets. Each superdroplet corresponds to a multiplicity $\xi_i(t)$ of "real" droplets which exist in the same gridbox and have identical such attributes.

The proposed breakup algorithm unifies the representation of collisional coalescence and breakup and builds on the original coalescence Monte Carlo steps in Shima et al. As in this original SDM, we begin by selecting pairs of superdroplets to consider collisions:

1. All superdroplets within a cell are randomly ordered in a list of non-overlapping pairs (j_α, k_α) where j and k are the superdroplet indices, and α refers to the pair index.

Next, we determine how many collisions, γ_α, occur for the pair α in the time step:

2. The probability of collision between droplets i and j within a small volume ΔV is given by

$$P_{i,j} = K_{i,j}\frac{\Delta t}{\Delta V}, \tag{2.1}$$

   where $K_{i,j}$ [$\mathrm{m}^3\ \mathrm{s}^{-1}$] is the volumetric rate of collisions based on the properties of droplets i and j, and Δt is the model time step. The scaled probability of collision $P_\alpha^{(s)}$ for this pair α accounts for the multiplicities of the colliding superdroplets:

$$P_\alpha^{(s)} = \max(\xi_j, \xi_k)P_\alpha. \tag{2.2}$$

   Only a subset $\lfloor n_s/2 \rfloor$ of possible SD pairs are considered out of all possible superdroplet pairs at each time step. Therefore, the probability is further scaled up to form the corrected probability of collision:

$$p_\alpha = \frac{n_s(n_s-1)}{2} / \left\lfloor \frac{n_s}{2} \right\rfloor P_\alpha^{(s)}. \tag{2.3}$$

The number of collisions that occur in this time step, γ_α, is then determined in a Monte Carlo step based on p_α. Taking $\phi_\alpha \in (0,1)$ to be a uniform random number,

$$\gamma_\alpha = \min(\lceil p_\alpha - \phi_\alpha \rceil, \lfloor \xi_{j_\alpha} / \xi_{k_\alpha} \rfloor). \tag{2.4}$$

Here, we assumed the superdroplets are ordered such that $\xi_{j_\alpha} \geq \xi_{k_\alpha}$. If $\gamma_\alpha = 0$, then no collisions occur.

The collision rate is then γ_α collisions per gridbox and per time step. Due to the constraint in equation 2.4 based on droplet multiplicity, some collisions which should occur probabilistically cannot if the donor superdroplet has insufficient multiplicity to collide p_α times. Therefore, a collision deficit $p_\alpha - \gamma_\alpha$ may be tracked as a tool to assess whether the model time step is sufficiently small (elimination of the the collision deficit is used for adaptive step size control in the SDM implementation used herein (Arabas et al., 2022)).

In the original SDM, particles coalesce as long as $\gamma_\alpha > 0$, as the rate of collisions is taken to refer only to collisional coalescence. However, when we consider collisional breakup, an additional Monte Carlo step must be taken to determine whether the particles coalesce or break up. This is determined based on a coalescence efficiency (or collection efficiency) E_c, which generally depends on properties of the colliding particles such as their fall speed, mass, and surface tension. We additionally account for the fact that in some collisions, droplets may bounce off of one another elastically by including an optional additional parameter for the breakup efficiency, E_b. This second Monte Carlo step is summarized as follows.

3. Compute the dynamic that occurs: coalescence, breakup, or bounce (nothing). A second uniform random number ϕ'_α determines the outcome:

$$\text{dynamic}_\alpha = \begin{cases} \text{coalescence}, & \phi'_\alpha \leq E_c(\mathbf{a}_j, \mathbf{a}_k) \\ \text{breakup}, & E_c(\mathbf{a}_j, \mathbf{a}_k) < \phi'_\alpha \leq E_b(\mathbf{a}_j, \mathbf{a}_k)(1 - E_c(\mathbf{a}_j, \mathbf{a}_k)) + E_c(\mathbf{a}_j, \mathbf{a}_k) \\ \text{bounce}, & \phi'_\alpha > E_b(\mathbf{a}_j, \mathbf{a}_k)(1 - E_c(\mathbf{a}_j, \mathbf{a}_k)) + E_c(\mathbf{a}_j, \mathbf{a}_k) \end{cases} \tag{2.5}$$

Once the dynamic is determined, a fragment size is sampled if necessary:

4. Sample a fragment size $m_{f,\alpha}$ (mass) from a fragment size distribution, $P_{f,\alpha}$, with cumulative distribution function (CDF) $C_{f,\alpha}(\phi)$ that depends on the colliding particle attributes. A related variable, $N_{f,\alpha}$, is taken to denote the number of fragments that would form in a collision between droplets of mass m_j and m_k: $N_{f,\alpha} = \frac{m_k+m_j}{m_{f,\alpha}}$.

Finally, updating of multiplicities and attributes proceeds based on the selected dynamic, number of collisions, and sampled fragment size (if applicable):

(a) For coalescence:

$$\begin{cases} \xi'_j = \xi_j - \gamma_\alpha \xi_k \\ \mathbf{a}'_j = \mathbf{a}_j \\ \mathbf{a}'_k = \mathbf{a}_k + \gamma_\alpha \mathbf{a}_j \\ \text{if } \xi'_j = 0, \text{ then } \xi'_j, \xi'_k = \xi_k/2, \ \mathbf{a}'_j = \mathbf{a}_k \end{cases} \tag{2.6}$$

γ_α collisions occur between the two superdroplets, thus the more multiplicitous superdroplet j transfers $\gamma_\alpha \xi_k$ "real" droplets to coalesce with the less multiplicitous superdroplet k. j thus maintains its previous attributes, while the attributes of k are recomputed as a multiplicity-average of its initial attributes and the transferred j attributes. The coalescence rate is incremented by $\gamma_\alpha \xi_k$. The final step in updating multiplicities and attributes serves to conserve the number of superdroplets with nonzero multiplicity in the simulation in the case that all droplets within superdroplet j are depleted. Unlike in Shima et al., 2009, where superdroplet j is discarded, this approach sets properties of superdroplet j to be identical to k, and both j and k to half of k's multiplicity to conserve mass.

(b) For breakup:

In some cases, only $\gamma_{jk} \leq \gamma_\alpha$ breakups can occur for a given superdroplet pair without encountering negative multiplicities. We compute this maximum possible number of breakup steps and update the superdroplet properties using a recurrence relation (assuming $\gamma_\alpha > 0$), and track a breakup deficit rate of $\gamma_\alpha - \gamma_{jk}$. (Alternatively, one may perform substepping of the breakup event.) The particle attributes are updated such to be consistent with the result of several breakup steps with $\gamma_\alpha = 1$ occurring in sequence, always producing fragments of size $m_{f,\alpha}$.

$$\begin{cases} \gamma_{jk} = 0 \\ \xi_j^{\text{transfer}} = 0, \quad \xi_{j,\text{next}}^{\text{transfer}} = \xi_k \\ \xi_k^{\text{new}} = \xi_k, \quad \xi_{k,\text{next}}^{\text{new}} = \xi_k \left(\frac{m_j + m_k}{m_{f,\alpha}}\right) \\ \text{while } \gamma_{jk} < \gamma_\alpha \text{ and } \xi_{j,\text{next}}^{\text{transfer}} \leq \xi_j : \\ \qquad \xi_j^{\text{transfer}} = \xi_{j,\text{next}}^{\text{transfer}} \\ \qquad \xi_k^{\text{new}} = \xi_{k,\text{next}}^{\text{new}} \\ \qquad \gamma_{jk} = \gamma_{jk} + 1 \\ \qquad \xi_{j,\text{next}}^{\text{transfer}} = \xi_{j,\text{next}}^{\text{transfer}} + \xi_{k,\text{next}}^{\text{new}} \\ \qquad \xi_{k,\text{next}}^{\text{new}} = \xi_{k,\text{next}}^{\text{new}} \left(\frac{m_j}{m_{f,\alpha}}\right) + \xi_{k,\text{next}}^{\text{new}} \end{cases} \tag{2.7}$$

Note that the transfer of multiplicity to superdroplet k begins with a multiplication by the number of fragments that would result from a real collision between droplets of size m_j and m_k. Because the fragment size is conserved in subsequent iterations, the additional multiplicity transfer comes only from the more multiplicitous superdroplet j which continues to transfer mass to k during collisions.

$$\begin{cases} \xi_j' = \xi_j - \xi_j^{\text{transfer}} \\ \xi_k' = \xi_k^{\text{new}} \\ \mathbf{a}_k' = \frac{\xi_k \mathbf{a}_k + \xi_j^{\text{transfer}} \mathbf{a}_j}{\xi_k^{\text{new}}} \\ \text{if } \xi_j' = 0, \text{ then } \xi_j', \xi_k' = \xi_k^{\text{new}}/2, \ \mathbf{a}_j' = \mathbf{a}_k' \end{cases} \tag{2.8}$$

The breakup rate is incremented by $\gamma_{jk}\xi_k$. The breakup deficit rate is incremented by $(\gamma_\alpha - \gamma_{jk})\xi_k$.

(c) For bounce:

No update is made to droplet multiplicities or attributes, and only the collision counter is incremented.

2.2.3 Additional Implementation Details

This method of breakup allows for the splitting of a coalesced transition state into a non-integer number of fragments ($N_{f,\alpha} = \frac{m_k+m_j}{m_{f,\alpha}}$ need not be integer), depending on the sampled fragment size. For instances where it may be desirable to preserve superdroplet multiplicities as integers, we recommend rescaling the multiplicities after the breakup step by a factor of $r_k = \lceil \xi_k \rceil / \xi_k$, and the multiplicities correspondingly by $1/r_k$ such that extensive attributes (including mass) are conserved.

The presence of a "breakup deficit" in the case where $\gamma_{jk} < \gamma_\alpha$ can be averted by substepping, though this is inadvisable for parallel applications of the SDM. Furthermore, superdroplet multiplicities may increase without bound according to the algorithm as presented above, which can lead to numerical artifacts and instability within a simulation. A set of limiters preventing runaway multiplicity is discussed in Appendix 2.6. Finally, a method for sampling a fragment size from a nonlinear empirical distribution, such as Straub 2010, is discussed in Appendix 2.6.1.

2.3 Validation and Convergence Properties

In order to validate the proposed Monte Carlo algorithm which encompasses both coalescence and breakup, we compare to an analytical solution to the generalized stochastic collection equation (SCE, or Smoluchowski equation). The approach is similar to that of Lee and Matsoukas, 2000 and compares to the solution of Srivastava, 1982, which uses constant-in-time and attribute-independent coagulation and breakup rates. Discussion of the solution outside of cloud-physics can be found, e.g., in Hansen (2018, eq. 8.58 therein) where it is presented in the context of polymerization-depolymerization modeling (see Blatz & Tobolsky, 1945, for a relevant seminal work featuring analytic SCE solutions). Specifically, herein we compare to the results including processes of binary coagulation and breakup only, neglecting spontaneous fragmentation.

The deterministic solution to the SCE relates the evolution in time of the ratio m of mean mass of particles to the fragment mass m_{frag} in the case of constant-coefficient binary coagulation and breakup is given by (notation as in eq. 13 in Srivastava, 1982):

$$m(\tau) = m(0)e^{-\beta_\star \tau} + \left(1 + \frac{1}{2\beta_\star}\right)\left(1 - e^{-\beta_\star \tau}\right). \tag{2.9}$$

Case	c (s^{-1})	β (s^{-1})	m_{frag} (g)
breakup-only	10^{-15}	10^{-9}	0.25
coalescence-only	0.5×10^{-6}	10^{-15}	-
coalescence-breakup	0.5×10^{-6}	10^{-9}	0.25

Table 2.1: Coalescence rate (c), breakup rate (β), and fragment mass values for different simulation setups. The inactive process rates are set to $10^{-15}(s^{-1})$ rather than exactly 0 in accordance with the solutions of Srivastava, 1982.

The solution is given using non-dimensional variables defined by

$$\tau = cMt \tag{2.10}$$

$$\beta_* = \beta/c, \tag{2.11}$$

where t is time, c is a constant coalescence rate, M is the ratio of total mass of the system to the fragment mass m_{frag} and β is a constant breakup rate. The rates c and β correspond to efficiencies $E_c = c/(c+\beta)$ and $E_b = 1$ (no bouncing) with the corresponding collision kernel $K = c + \beta$. Of note, the solution does not depend on the initial particle size spectrum — only on the initial mean mass $m(0)$.

Two caveats are involved in comparing SDM results against SCE solutions in this test set up. First, the constant collision kernel admits collisions of same-sized particles, whereas collisions with a single superdroplet are not included in the present SDM implementation. This discrepancy should diminish with increasing numbers of super-particles (and hence decreasing values of multiplicities) down to zero for a one-to-one simulation with multiplicities of unity. The second caveat involves the assumption of breakup resulting in only a single fragment size, m_{frag}. This simplification is required to attain the analytic solutions of Srivastava, 1982, but removes the final of four stochastic elements in the proposed superdroplet breakup algorithm. As such, the following comparisons are useful to understand the convergence properties of the proposed algorithm's first three stochastic elements (superdroplet sampling, collision probability, coalescence probability) only.

We analyze three comparison cases in a zero-dimensional box setting: breakup only ($c \approx 0$), coalescence only ($\beta \approx 0$), and coalescence plus breakup. The parameter values defining the three aforementioned cases are summarized in Table 2.1. These simulations utilize the open-source Pythonic superdroplet code 'PySDM' (Bartman, Bulenok, et al., 2022; de Jong et al., 2023). For each test case, we perform the simulation at a few resolutions (number of superdroplets). Note while the SDM

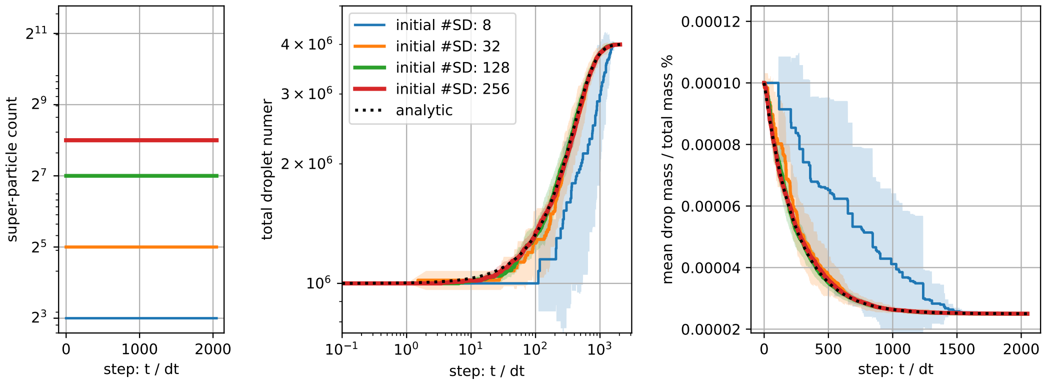


Figure 2.3: Mean (solid line) and standard deviation (shading) for the time evolution of the breakup-only dynamics, including the analytical solution of Srivastava, 1982 and the SDM using 8, 32, 128, or 256 superdroplets.

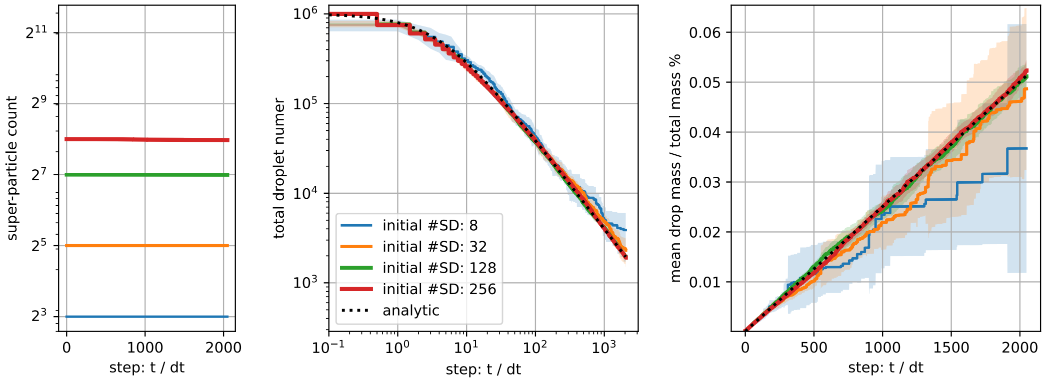


Figure 2.4: Mean (solid line) and standard deviation (shading) for the time evolution of the coalescence-only dynamics, including the analytical solution of Srivastava, 1982 and the SDM using 8, 32, 128, or 256 superdroplets.

implementation guarantees that the number of super-particles cannot increase, a superdroplet may be removed from the system during a collision event (for details see point (5) in section 5.1.3 in Shima et al., 2009). The simulations are performed for 2048s with 1s timesteps with adaptive collision substepping enabled. The initial size distribution is monodisperse with equal multiplicities for all superdroplets. These settings correspond to a population of 10^6 particles in one cubic metre, with each droplet having initial mass of 1g. In the solution of Srivastava (1982), it is assumed that all fragments resulting from breakup are of equal size. Here, the constant fragment mass is set (arbitrarily) to 0.25g.

Figures 2.3–2.5 present the mean and standard deviation of 10 realizations of the SDM simulations, versus the analytic solutions. Both the breakup-only (Figure 2.3)

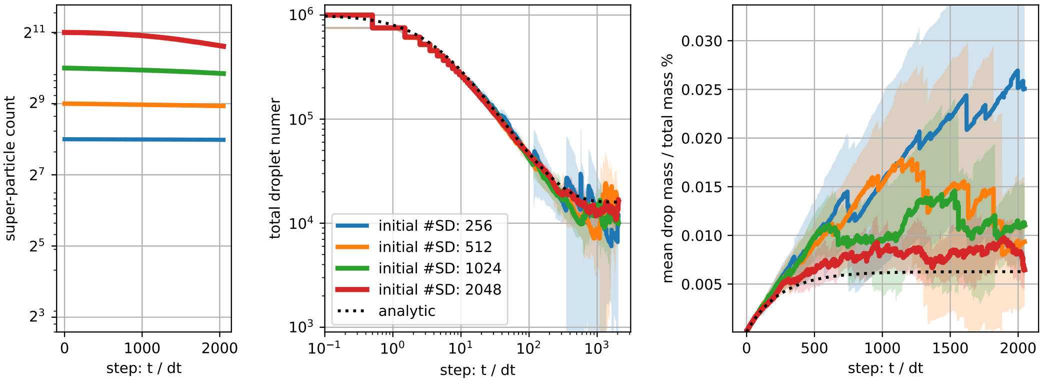


Figure 2.5: Mean (solid line) and standard deviation (shading) for the time evolution of the coalescence-breakup combined dynamics, including the analytical solution of Srivastava, 1982 and the SDM using 256, 512, 1024, or 2048 superdroplets.

and breakup-plus-coalescence (Figure 2.5) cases feature asymptotic values of droplet number and mass, corresponding to all droplets having the size m_{frag} (breakup-only) or a balance between fragmentation and coalescence of droplets (breakup-plus-coalescence). In all three sets of dynamics, increasing the number of superdroplets robustly reduces the ensemble spread and improves the match with the analytic solution. Much higher resolution (on the order of thousands superparticles) is required to match the analytic solution for the breakup-coalescence equilibrium than for simulations featuring each process separately. This discrepancy is related to the additional stochastic step of selecting the breakup or coalescence dynamic. Extension of this logic implies that even more superparticles would be required to match the exact asymptotic behavior of a system with a distribution of fragment sizes, due to yet an additional stochastic element of sampling.

In Figure 2.5, the simulation with the largest number of superdroplets displays removal of superdroplets from the system. The removal can happen when coalescence leads to zero multiplicity in one of the resultant superdroplets, which then cannot be split into two superdroplets. This is more likely when multiplicities are on average lower (i.e. when more superdroplets are used). In the limit of few superdroplets, both the breakup-only and the coalescence-only simulations are characterized by underestimation of the process rate, which is consistent with the SDM implementation that neglects collisional dynamics within a superdroplet.

2.4 Numerical Experiments and Discussion

To demonstrate the behavior of a particle population under the proposed breakup algorithm, we focus here on sensitivity studies in a zero-dimensional box setting and in a one-dimensional rainshaft setup. The coalescence efficiency and fragment size distribution are varied across the simulations while assuming a consistent initial PSD and collision rate. We consider both artificial parameterizations that assume fixed values for the efficiency and FSD, as well as empirical parameterizations for comparison. These selections are made to highlight the variability in results that is due to the additional stochastic components of the breakup algorithm while maintaining consistency across other choices such as the initial distribution, rate of collisions, and aerosol and updraft parameters (in the case of 1D test cases). No resampling of superdroplets is performed throughout the simulation.

The simulations presented in this section assume a canonical geometric collision kernel, where the volumetric rate of collisions K_{jk} [m^3 s^{-1}] between droplets with the properties of superdroplets j and k is given by

$$K_{jk} = \pi(R_j + R_k)^2 |v_j - v_k|, \tag{2.12}$$

where R_j is the radius of particle j and v_j is the terminal velocity/fall speed of particle j, computed using the parameterization of Gunn and Kinzer, 1949. This kernel assumes that all collisions result from interception of particles with different fall speeds, neglecting non-vertical motions, turbulence, and impacts of particles on the local flow field. As in Shima et al., 2009, collisions within a superdroplet (i.e. collisions between droplets represented by the same superdroplet) are neglected — in line with the assumption of a geometric collision kernel which precludes collisions between equally-sized droplets, as they have the same terminal velocity. (These intra-superdroplet collisions must be considered when using a turbulent collision kernel parameterization.)

2.4.1 Particle Size Distribution

The zero-dimensional box simulations include collisional-coalescence and collisional-breakup dynamics only. The droplet size distribution is initialized to an exponential distribution in mass x, given by $n(x) = x_0 \exp(-x/x_0)$ with the characteristic size $x_0 = (4\pi/3)R_0^3$ set using $R_0 = 30.531\mu$m as in Shima et al. (2009). The simulations employ $2^{13} = 8192$ superdroplets to represent a number density of 100cm^{-3} in a box of volume 1m^3 with a fixed time step of 1s. This choice of superdroplet quantity is sufficient to produce consistent results in the PSD across realizations using a different random seed, and was shown by Shima et al. to closely match the exact PSD in a similar box model simulation of collisional coalescence.

The coalescence efficiency is chosen to be either a constant value (for sensitivity studies), or the empirical coalescence efficiency of Straub et al., 2010 which depends on the Weber number of the colliding droplet pair. (The Weber number is a ratio of kinetic collisional energy and surface tension, and relates to the stability of a droplet pair under collision.) We consider three types of fragmentation functions: (1) a constant fragment number N_f, in which the particle-size distribution (PSD) is a delta function $P_f(m_{f,\alpha}) = \delta(m_{f,\alpha} - \frac{m_j+m_k}{N_f})$; (2) an exponential distribution $P_f(m_{f,\alpha}) \sim \exp(-m_f/\mu)$ where the scale μ is specified; and (3) the empirically derived fragmentation function of Straub et al., which uses four modes of fragmentation represented by lognormal or normal subdistributions.

Particle size distributions are displayed as the number distribution or as the marginal mass dm distribution in particle radius R, $g(R) = \frac{dm}{dln(R(x))} = \rho_l 3x^2 n(x)$ where $n(x)$ is the particle size distribution in particle volume x, and ρ_l is the bulk density of liquid water. This metric, with units [kg m^{-3}], is defined consistently via the total mass density of particles in the system M:

$$M = \int_{-\infty}^{\infty} g(R) dln(R) = \int_{0}^{\infty} \rho_l x n(x) dx.$$

The marginal mass distribution is computed by binning the resulting superdroplets into 128 logarithmically-spaced size bins between particle radius 1µm and radius 10mm. We separate the simulations into those which use a deterministic fragmentation function, in which breakups result in a constant number of fragments in any given collision; a stochastic fragmentation function with fragment sizes sampled from a specific distribution; and a size-dependent fragmentation function, where the fragment sizes are sampled from a distribution whose parameters depend on the colliding particles. We further include experiments exploring the use of a fixed coalescence efficiency versus a particle-attribute-derived coalescence efficiency. This separation elucidates which aspects of the particle population behavior are attributable to stochastic sampling of the fragmentation function, or related to particle-property-dependent parameters such as Weber number.

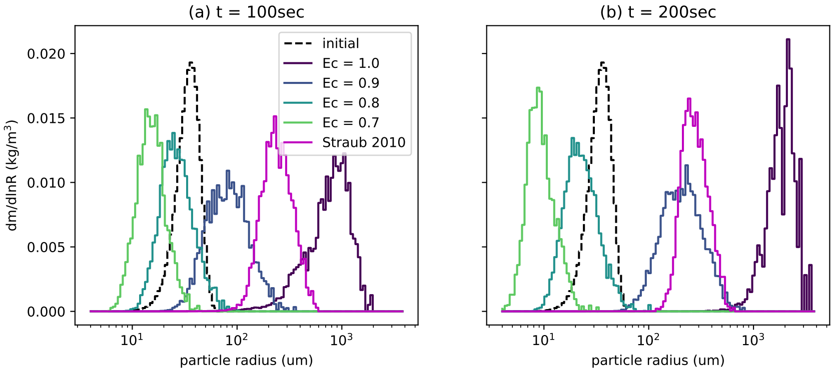


Figure 2.6: Particle size distribution with varying coalescence efficiencies under a geometric collision kernel after 100s (left) and 200s (right). The breakup fragmentation function is deterministic, with the fragment size determined as 1/8th the sum of the colliding droplet sizes. The dashed black line represents the initial PSD, and solid lines represent various fixed values of the coalescence efficiency. The pink line corresponds to a size-dependent coalescence efficiency from Straub et al.

2.4.1.1 Sensitivity Studies: Deterministic and Size Independent Fragmentation

First we investigate the sensitivity of the PSD evolution to the coalescence efficiency, using four values of a constant-valued efficiency E_c between 0.7 and 1.0 ($E_c = 1.0$ corresponds to coalescence-only) and a particle-size dependent E_c parameterization (Straub et al., 2010). All simulations use a deterministic fragmentation function in which all single-step collisional breakups result in $N_f = 8$ fragments:

$$P_{f,\alpha}(m_f) = \delta\left(m_f - \frac{m_j + m_k}{N_f}\right). \tag{2.13}$$

Figure 2.6 displays two snapshots of the PSD under this set of dynamics, demonstrating the additional Monte Carlo step of selecting whether coalescence or breakup occurs, independent of sampling a fragment size. As expected, the initial PSD broadens and shifts toward larger droplets at 100s, with the largest values of fixed E_c leading to the largest increase in average particle mass. However, after 200s, the PSD for the $E_c = 0.8$ case remains approximately steady with a mean size that is smaller than the initial distribution mean, demonstrating that coalescence and breakup are approximately balanced in this case.

By contrast, the PSD for the Straub 2010 parameterization of E_c initially between the $E_c = 0.9$ and $E_{c}. = 1.0$ simulations at 100s, but narrows without shifting toward much larger droplets after further time has elapsed, leading to a dominant mode

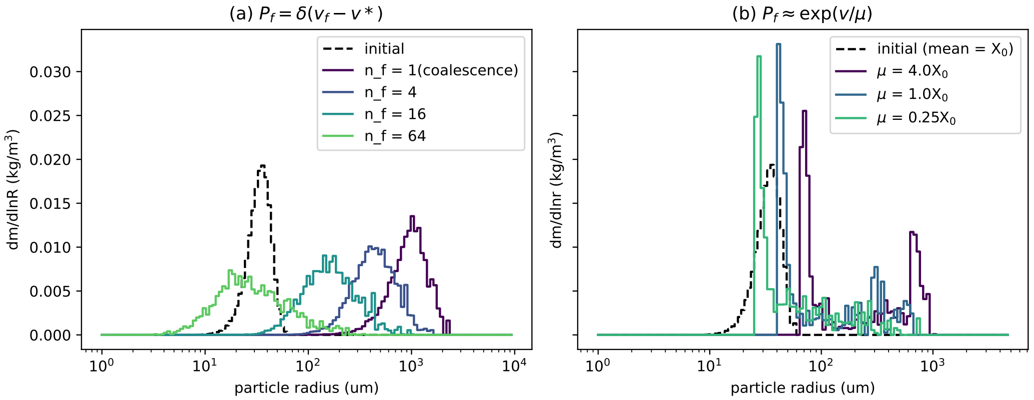


Figure 2.7: Sensitivity to fragmentation function of PSDs following collisions with a geometric kernel and fixed coalescence efficiency of $E_c = 0.95$ in a deterministic and stochastic fragmentation function case. (Left) The fragment size is fixed by a divisor of the sum of colliding particle volumes; (right) fragment size is sampled from an exponential distribution with varying means μ determined as a multiple of the initial distribution mean. The initial distribution is shown as a black dashed line in each figure.

that is more similar to the $E_c = 0.9$ case. This empirical parameterization also shows evidence of approaching a steady state distribution, in which coalescence and breakup rates are matched on average, driving the PSD to a stationary state. The Straub 2010 parameterization decreases exponentially with the colliding particle Weber number, which is correlated with the size and relative terminal velocity of the colliding particles. Two colliding particles of comparable size have a low relative terminal velocity, therefore as the PSD shifts toward larger coalesced droplets, there is a competing effect between a larger particle size increasing the Weber number, and decreased relative terminal velocity reducing it. This competition produces the stationary behavior and narrowing of the PSD observed in this case.

Next we consider the PSD evolution when the coalescence efficiency is held fixed at a constant value and the fragmentation function is varied. In Figure 2.7(a), we consider a deterministic fragmentation function where the number of fragments from a single breakup event is fixed (as in Figure 2.6), as well as an exponential fragment size distribution with scale μ specified as a multiple of the initial mean particle mass x_0. When the number of fragments is fixed, results using the largest number of fragments display the smallest mean particle size and broadest spectra. The first behavior is expected, as a larger value of N_f results in smaller typical fragment sizes. The broadening of the spectrum can be attributed to a wider range

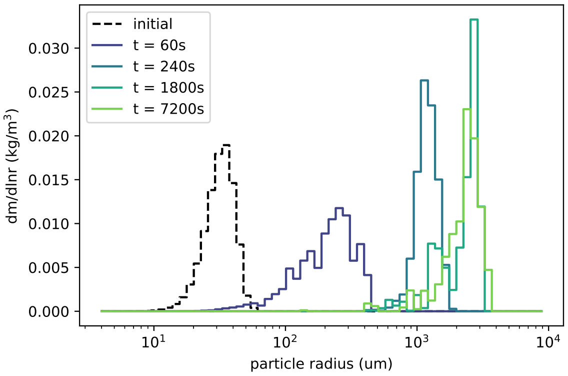


Figure 2.8: Initial PSD (black dashed) and PSD's following collisions with a geometric kernel, the Straub 2010 collection efficiency, and the Straub 2010 fragmentation function (Straub et al., 2010) after several elapsed times (colors).

of collision rates between small droplets (which result from fragmentation), and is generally an expected outcome of including collisional breakup.

When the fragment size is sampled from an exponential distribution (Figure 2.7(b)), the resulting spectra are bimodal, with a large-droplet mode that larger for larger choices of the mean fragment size μ, and a narrow small-droplet mode that likewise depends on μ. The appearance of a second mode occurs when the fragment size is sampled from the left tail of the fragment size distribution, whereas the large mode corresponds to a droplets undergoing coalescence only, as in the $N_f = 1$ case. This behavior indicates that through stochastic sampling of the dynamic and fragment size together, the droplet population splits into one mode which fragments into smaller droplets, and a second mode which primarily undergoes coalescence and grows in size. Because larger droplets collide at much quicker rates than small droplets, the fragmented mode is less likely to collide and re-coalesce to form medium-sized droplets, while the coalesced-mode retains some probability of colliding and either growing (coalescing), or breaking up into smaller droplets. Thus we observe that the small-droplet-mode grows in this instance, with particles effectively become "stuck" in this dynamical regime due to the separation of scales in collision rates.

2.4.1.2 Steady State under Stochastic Size-Dependent Fragmentation

Finally, we consider an empirically derived coalescence efficiency and corresponding fragmentation function whose parameters depend on the colliding droplet properties (Straub et al., 2010). In Figure 2.8, we consider the evolution of the PSD under the Straub 2010 efficiency and fragmentation dynamics, beginning from the

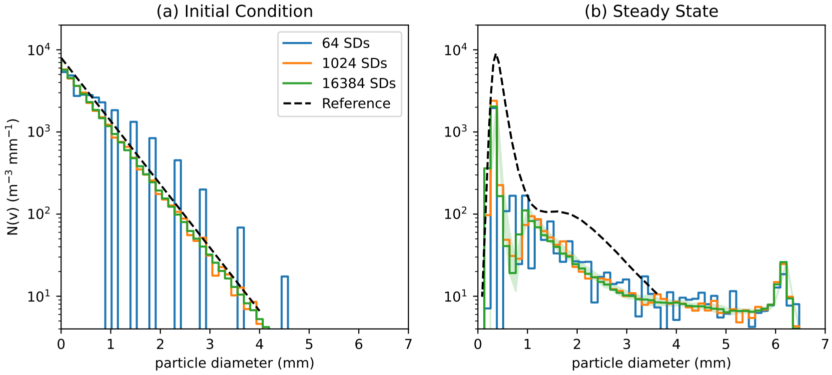


Figure 2.9: Initial PSD (left) and steady-state PSD (right) under collisions with the Straub 2010 parameterizations, including reproduction of the findings in figure 10 from Straub et al., 2010 (black dashed line), and simulations including the proposed breakup algorithm with different numbers of superdroplets. Steady state results are displayed as the mean plus or minus one standard deviation (shading) of 10 simulation instances.

same initial distribution as previous experiments. At first, the PSD broadens and shifts towards larger droplet sizes, as in Figure 2.6, but the PSD after 7200s shows little difference from the PSD at 1800s. These results indicate the stationarity of the particle size distribution after sufficient time has elapsed: coalescence and breakup are balanced, as in the previous example. Contrasted with Figure 2.6, which used a deterministic size-independent fragmentation function, the stationary PSD resulting from the Straub et al. (2010) parameterization of fragmentation is less symmetric and contains multiple small peaks. This difference reflects the sampling from a multimodal distribution of fragment sizes, contrasting with the symmetric PSDs found from using a fixed number of fragments.

This empirical parameterization provides an additional opportunity for validation of the breakup algorithm on top of the analytical results presented in Section 2.3. Figure 2.9 compares results of this SDM implementation against figure 10 of Straub et al., 2010 (note that we now plot number concentration rather than mass concentration with a logarithmic y-scale and linear x-scale, as in Straub et al., 2010 and McFarquhar, 2004). Particles are initialized as a Marshall Palmer distribution (exponential in droplet diameter) with rain rate 54mmhr^{-1}; superdroplet sizes are sampled logarithmically over this diameter range, rather than using constant multiplicity. With only 64 superdroplets, the SDM struggles to adequately sample the low concentration but important rain droplets with diameters exceeding 1 mm, while using over 1000 superdroplets is more than sufficient. The simulation is run using a 1s time step for 7200s to approximate a steady state in Figure 2.9b. The SDM appears to converge using between 1000 to 16,000 superdroplets, with more superdroplets being necessary to reduce spread of the results for small particle diameters. The SDM approach captures the size and amplitude of the first mode of the steady state distribution reasonably well, but shows some discrepancies at at intermediate particle sizes, predicting a secondary mode at 1.2mm diameter rather than the less pronounced shoulder at 2mm. Inspecting the resulting distributions at large particle sizes reveals a larger quantity of 6-7mm diameter particles in the SDM simulations, which are not seen in the work of Straub et al., 2010. This behavior indicates that some large superdroplets coalesce rather than breaking up, which may be reflective of the multiplicity-limiter in the model which performs coalescence rather than breakups which would result in a large increase in multiplicity (i.e. sampling a small fragment size during collision involving one or more large droplets). The coalescence efficiency tends toward zero for any droplet larger than 6 mm in diameter, which explains why the large droplet mass is concentrated at this size. In a realistic setting, however, most rain-range droplets would be expected to sediment before attaining this large 6mm size through coalescence. Indeed, later results presented in Figure 2.11 do not show significant mass concentration at droplet sizes larger than 1 mm radius, indicating that this error in the steady state does not present in the transient rainshaft simulation.

2.4.2 Cloud and Precipitation Properties

Next we consider the impact of collisional breakup in a one-dimensional warm rain setting that includes condensation/evaporation (including aerosol activation/deactivation), collisions, and transport of particles within the column through advection and sedimentation/precipitation. These 1D simulations are based on the kinematic framework of Shipway and Hill, using a fixed profile of dry-air potential temperature and dry-air density $\rho_d(z)$, and a resolved budget of water vapor (advection and coupling with vapor uptake and release by particles). Ventilation effects are neglected, therefore rain evaporation is underpredicted and most rain sediments and leaves the simulation domain. The vapor advection is solved using the MPDATA algorithm on a columnar grid with vertical spacing of 100m (employing the PyMPDATA implementation Bartman, Banaśkiewicz, et al., 2022). An aerosol population with hygroscopicity $\kappa = 0.9$ is initialized throughout the vertical domain with $2^8 = 256$ superdroplets per gridbox. This choice of 256 superdroplets per gridbox reflects the higher computational demands of the one dimensional simulation compared to the box model and still produces statistically convergent results in the macroscopic quantities investigated. For the first 600s of spin-up, condensation-evaporation (including aerosol activation using kappa-Köhler theory; implementation of these processes follows that of Arabas et al., 2015) and particle advection with the specific updraft are the only active dynamics, with a time-varying updraft momentum flux of $\rho_d w(t) = 6\text{kgm}^{-3}\text{ms}^{-1} \sin(\pi t/600\text{s})$. After this spin-up time, the updraft velocity is set to 0, and the processes of particle displacement due to sedimentation and collision-coalescence-breakup begin. The time step is fixed at 5s throughout the simulation.

The test cases demonstrated here include a no-breakup case, a property-independent breakup case where the coalescence efficiency is fixed and fragment sizes are sampled from a fixed distribution, and the particle-property-dependent empirical coalescence efficiency and fragmentation parameterizations from Straub et al., 2010. All simulation parameters including the thermodynamic constants, number of superdroplets, geometric collision rate (equation 2.12),and the Gunn and Kinzer terminal velocity parameterization are fixed, while the coalescence efficiency and FSD are varied. In the no-breakup case, all collisions result in coalescence. In the property-independent breakup case, we fix $E_c = 0.95$ for all superdroplet collisions based on the correspondence in Figure 2.6 to the empirical coalescence efficiency. This case samples fragment sizes from a Gaussian distribution in particle volume with mean radius 30µm and standard deviation 15µm. In contrast to the property-independent case, in which the fragmentation parameters are fixed regardless of the colliding droplets, the property-dependent setting is based on empirical evidence, and is expected to be more reflective of the variability of real clouds. In both the property-independent and -dependent cases, the breakup efficiency is set to $E_b = 1$ such that all collisions result in either coalescence or breakup. To contrast the behavior of the three cases, we consider the hydrometeor population at various altitudes throughout the simulation, as well as collision process rates and aerosol processing rates.

2.4.2.1 Hydrometeor and Cloud Quantities

The mixing ratio of cloud droplets (activated droplets of no more than 50µm radius), rain droplets (radius greater than 50µm), and the number concentration of unactivated aerosols are displayed for the three test cases in Figure 2.10. The no-breakup simulation forms a cloud due to activation of aerosols between 600m and 3800m altitude until the updraft is terminated after 600s. Larger rain-range droplets form

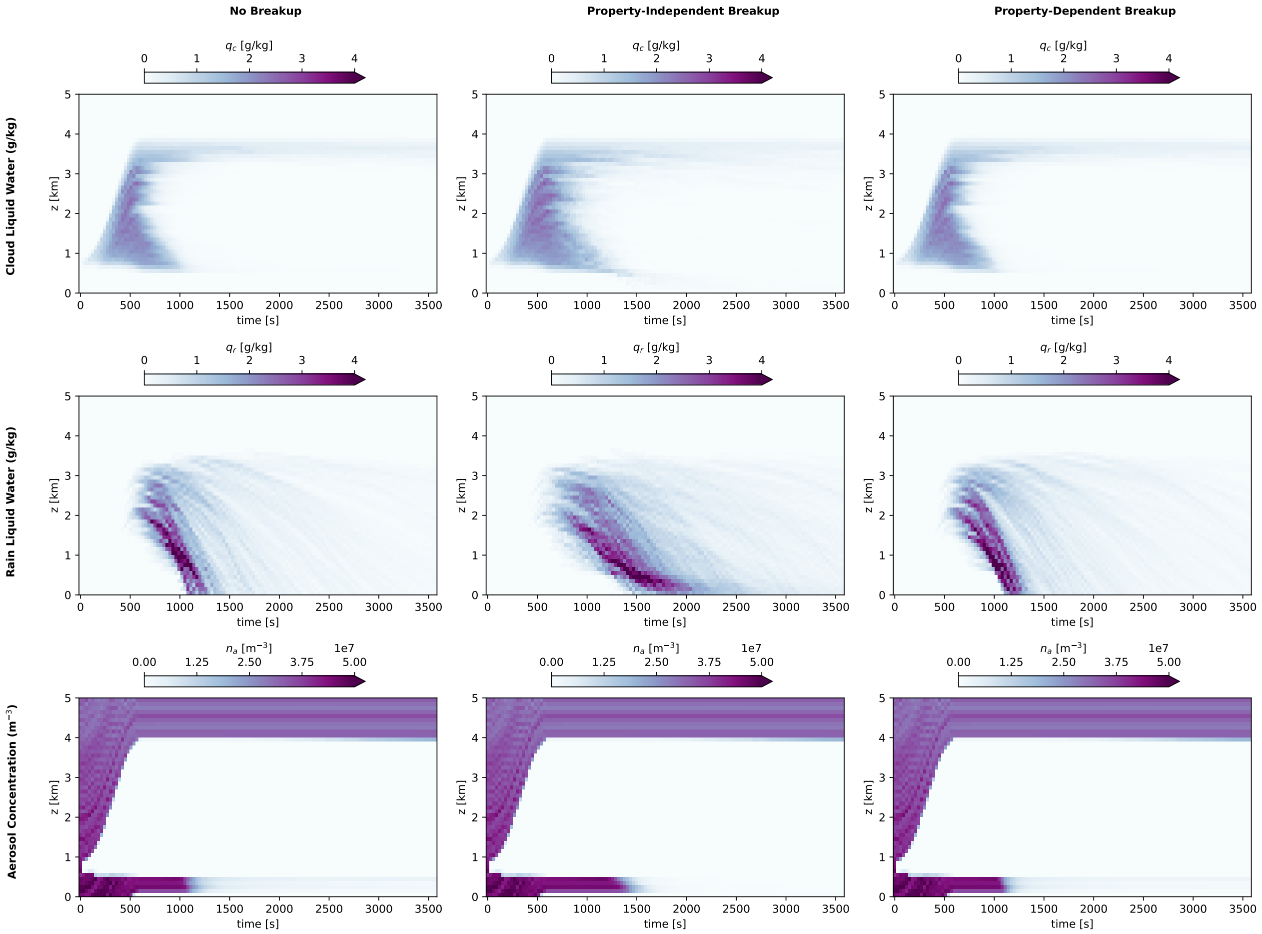


Figure 2.10: Hydrometeor concentrations without breakup (left column), with breakup using a property-independent coalescence efficiency (middle column), and with breakup following the property-dependent Straub et al. (2010) parameterizations (right column). Included are cloud water mixing ratio (top row), rain water mixing ratio (middle row), and aerosol number concentration (bottom row).

from collisional coalescence and begin to sediment out of the system in clusters, visible as distinct streaks in the (t, z) plane, with surface precipitation beginning around 1100s into the simulation, depleting the cloud droplet population. The vertically-averaged particle size spectra in Figure 2.11 demonstrate these qualitative changes as well. At 600s, the majority of particles are micron-sized aerosols with cloud droplets beginning to form. At later times of 900s and 1200s, the particles rapidly grow to tens or hundreds of microns in size, and at 1800s, the mass distribution in the no-breakup case shows significant depletion due to precipitation of large particles.

When property-independent breakup is included, a higher concentration of cloud-

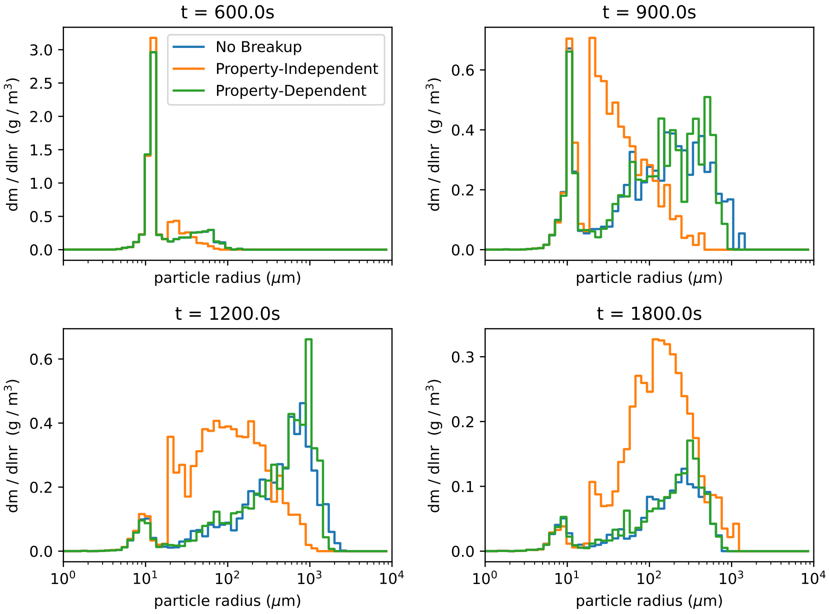


Figure 2.11: Vertically-averaged particle size spectra (as a marginal mass distribution) at four times selected from the 1D rainshaft simulations for each set of dynamics (colors).

sized droplets persists at cloud base, and the surface precipitation is delayed and spread out relative to the case with no breakup. This behavior indicates that rain droplets favorably break up within the cloud and especially near cloud base, fragmenting into smaller cloud droplets (the mean of the fragment size distribution is 30µm radius, only slightly lower than the rain size range) with a lower sedimentation rate. We observe this behavior in Figure 2.11 as well: the property-independent case shows much smaller average particle sizes at 900 and 1200s, as well as more mass remaining in the system at 1800s due to the delayed precipitation. Furthermore, the aerosol population below cloud base is not depleted as quickly in this property-independent case, indicating a reduction in aerosol scavenging and washout that is consistent with the lower precipitation rates. These phenomena are consistent with documented impacts of collisional breakup such as reduced surface precipitation (Seifert et al., 2005), and show that the proposed algorithm can meaningfully represent the breakup process.

The empirical property-dependent breakup case using the Straub et al. parameterizations displays hydrometeor populations that are more similar to the no-breakup case, indicating that the choice of $E_c = 0.95$ in the property-independent case likely overestimates the rate of collisional breakup when condensation and evaporation are present (contrasted with Figure 2.6). As in the no-breakup case, the property-dependent empirical case displays distinct streaks of precipitation, with surface precipitation initiated around 1100s. While the hydrometeor populations show only slight differences between no-breakup and property-dependent breakup in Figure 2.10, the size spectra at 900s in Figure 2.11 shows a somewhat narrower size distribution for the property-dependent case as particles approach the rain size range, suggesting the role of collisional breakup.

The relatively short updraft time and simple one-dimensional representation of this setup produce a short-lived cloud that is precipitating for only a few minutes. The likelihood of breakup in the Straub parameterization is strongly correlated with the size of the colliding droplets, with E_c approaching one for colliding droplets smaller than 1mm diameter, therefore we expect to see a stronger impact of including SDM breakup in a strongly precipitating convective case. Due to the complexity and feedbacks inherent to representing a superdroplet-coupled flow field as well as mixed-phase processes, such experiment is beyond the scope of the present work focused on the algorithm formulation. In deeper mixed-phase clouds, however, secondary ice production via ice-ice and ice-supercooled-liquid collisions are two analogous processes of collisional breakup that may be important (Hallett & Mossop, 1974; Harris-Hobbs & Cooper, 1987; James et al., 2021; Zhao & Liu, 2022). Multiphase superdroplet representation of these mechanisms will face a similar challenge of representing many different fragment sizes, thus the SDM breakup representation presented in this work could be extended to collisional ice processes in future research on secondary ice production and mixed-phase clouds.

2.4.2.2 Process Rates

Figure 2.12 displays the local rates of superdroplet collision (scaled by multiplicity), as well as distinguishing between rates of coalescence and rates of breakup. We see an expected correlation between the time and location of collisions in all three cases with the location of hydrometeors (outlined in black for cloud and red for rain)—as expected, a higher concentration of hydrometeors, particularly large rain-range hydrometeors corresponds to higher rates of all collisional dynamics. The rate of collisions increases throughout the simulation time, particularly near cloud base where the largest droplets are sedimenting and colliding at higher rates. The property-independent case is consistent with the other cases in displaying higher collision rates at cloud base, even though the droplets in this region are slightly

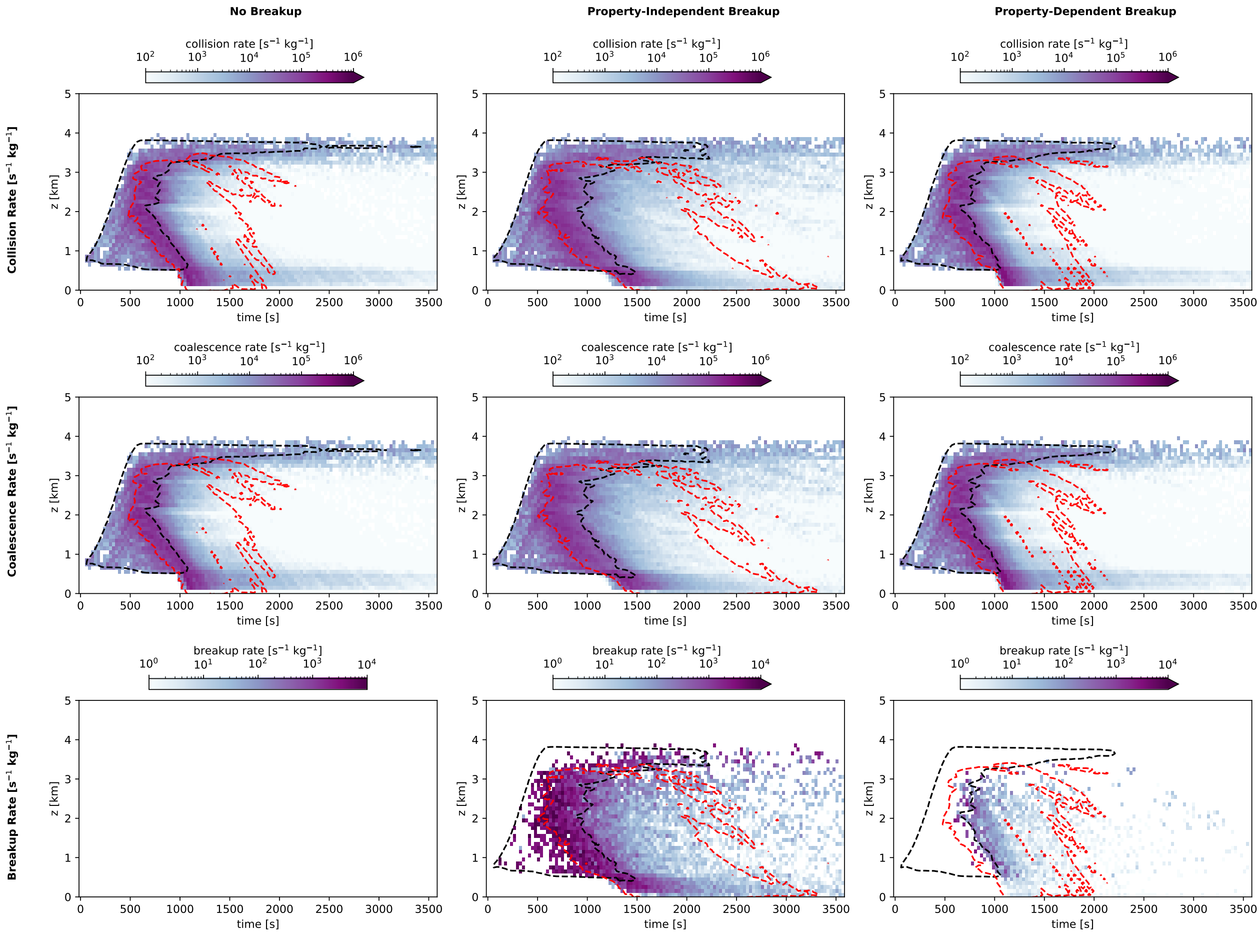


Figure 2.12: Collisional dynamic rates for the 1-dimensional case with (left) no breakup, (middle) breakup with a fixed coalescence efficiency, and (right) breakup with the Straub 2010 parameterizations. The dynamics shown include (top to bottom): collision rate, coalescence rate, and breakup rate. Dashed contour lines represent the level of $q_c = 0.4$g/kg (black) and $q_r = 0.4$g/kg (red), representing a cloudy and rainy region of the time-space domain, respectively.

smaller and fall in the cloud category.

All three cases display similar rates of collision and coalescence, with the highest of these rates occurring in the cloud among rain droplets, and below cloud base among precipitating rain droplets. In the no-breakup case, every feasible collision results in a coalescence, and the breakup rate is zero. When property-independent breakup is included, the time-space distribution of the breakup rate is nearly identical to that of the collision rate. This trend is consistent with the use of a uniform coalescence efficiency $E_c = 0.95$, which is agnostic to the size of the colliding particles. In contrast, the empirical property-dependent case sees collisional breakup primarily where larger rain droplets are present, consistent with the Straub et al. parameterization based on Weber number. Breakup events drop off quickly after the initial depletion of the cloud-sized droplets, as the coalescence efficiency for two large particles is much smaller than that of a small particle colliding with a large particle. As seen in Figure 2.11, breakup plays less of a role in determining the hydrometeor populations after 1200s of simulation time. Thus, the impacts of collisional breakup are limited in time and space to where large concentration of hydrometeors of both cloud and rain size are colliding. These results demonstrate that the SDM breakup algorithm can produce expected process behavior in both a property-independent setting, where the collision dynamics result in strong breakup, and in an empirically parameterized setting.

As noted in the discussion of Figure 2.10, the property-independent breakup case experiences a persistent population of aerosols below cloud base until 1500s, while the no-breakup and property-dependent cases demonstrate washout upon the earlier onset of precipitation. Collisional breakup could potentially introduce cloud droplets so small that they deactivate in their environment. In Figure 2.13, we investigate whether collisional breakup can induce significant changes to aerosol processing rates, considering aerosol activation, deactivation, and ripening.

The property-independent and no-breakup cases have nearly identical behavior in aerosol processing, consistent with the correspondence between their hydrometeor concentrations and collision process rates. In all three cases, a few superdroplets at cloud top encounter humidity close to their critical supersaturation, which results in the "ripening" processes of fluctuation between an activated and deactivated state due to competition when the supersaturation is insufficient to activate all aerosols (e.g., Arenberg, 1939; Wood et al., 2002). (We define ripening rate as the number of activated droplets growing through condensation per unit of time within a grid cell

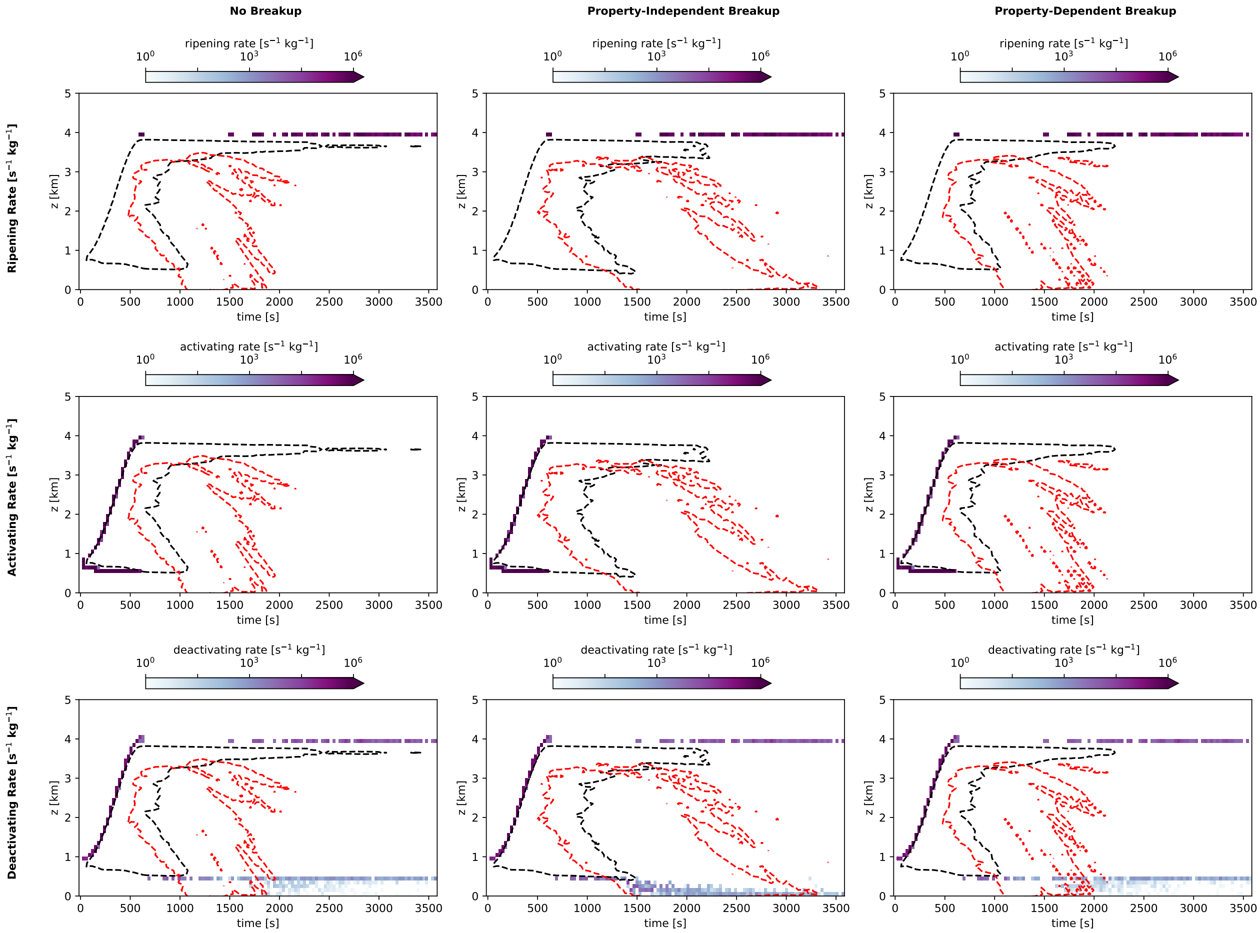


Figure 2.13: Aerosol processing rates for (left to right) no breakup, property-independent breakup, and Straub et al. (2010) parameterizations. Included are (top to bottom): ripening rate, activating rate, and deactivating rate. Dashed contour lines represent the level of $q_c = 0.4$g/kg (black) and $q_r = 0.4$g/kg (red), representing a cloudy and rainy region of the time-space domain, respectively.

in which deactivation simultaneously occurs on other particles.) Aerosols activate primarily at the start of the simulation when an updraft is present, defining the altitude boundaries of the cloud. No additional activation is seen in either instance including collisional breakup. Deactivation occurs among a few aerosols which activate and then rise in altitude beyond cloud top initially, and more strongly below cloud base as droplets sediment out of the cloud and evaporate. The property-independent case experiences much stronger deactivation at cloud base, which corresponds to the higher rate of fragmentation of droplets at this altitude. The no-breakup and property-dependent breakup cases display continued deactivation of aerosols at the cloud base height throughout the simulation, while the property-independent case shows only near-surface deactivation, suggesting aerosol scavenging by rain droplets throughout the more sustained precipitation process. These results indicate that collisional breakup could be a relevant process for future studies of aerosol-cloud effects, particularly in deeper-convective cases where collision rates are likely to be higher.

2.5 Conclusions

This work presents a superdroplet algorithm for collisional breakup that is scalable in avoiding creation of new superdroplets and physical in its ability to produce results in a box and one-dimensional setting that are consistent with the expected reduction and delay in rain formation. Furthermore, the algorithm produces hydrometeor populations and process rates that differ between a property-independent approach (with a fixed coalescence efficiency and fixed fragment size distribution), and a property-dependent approach using empirical parameterizations. These differences indicate the importance of random, stochastic events in warm rain microphysics, a trait which has also been documented in other microphysical phenomena such as giant CCN (Feingold et al., 1999; Yin et al., 2000). Without a scalable representation, the superdroplet method has heretofore been unable to capture these additional stochastic impacts of breakup, nor has it been applied to compare empirical parameterizations of coalescence and breakup, which contribute uncertainties to operational process, weather, and climate models.

This work provides the basis for a more complete representation of microphysical processes in particle-based simulations. For instance, a superdroplet representation that considers ice-phase hydrometeors and properties could probe processes of secondary ice production such as fragmentation and freezing of supercooled water droplets upon collision with ice (James et al., 2021), rime splintering (Hallett & Mossop, 1974), or ice-ice collisions (Harris-Hobbs & Cooper, 1987; Phillips et al., 2017). These and other mixed-phase processes are poorly understood due to challenges in obtaining direct observational or laboratory measurements, thus a high-fidelity particle-based representation such as the superdroplet method provides an ideal means for studying these phenomena. While the collisional breakup representation presented here does not address underlying uncertainties in parameterization of processes such as collision rates and phase change, it provides a path forward for more rigorous and complete studies of cloud microphysics.

Code and Data Availability

Implementation of this breakup algorithm in the SDM is available at https://doi.org/10.5281/zenodo.7851352. The simulations presented in this work (and all necessary input information) are available in the folder ’deJong_Mackay_2022’ at https://doi.org/10.5281/zenodo.7851288. The notebooks in this folder reproduce all results and figures presented in this study, with no external datasets required. The scripts run the relevant model configuration in a matter of minutes and plot the resulting output. All results presented in this paper can be reproduced by one of two means: (1) downloading and installing ’PySDM’ and ’PySDM-examples’ (e.g. using ’pip install’), and running the notebooks locally; (2) accessing the PySDM-examples repository online and running the examples notebooks in the folder ‘deJong_Mackay_2022’ on Google Colab. These codes, PySDM and PySDM-examples, are continuously under development at https://github.com/atmos-cloud-sim-uj/PySDM and https://github.com/atmos-cloud-sim-uj/PySDM-examples, and are further documented in a software publication (de Jong et al., 2023).

2.6 Appendix A: Limiters

In implementing collisional breakup for superdroplets, we suggest imposing a few limiters to enforce physical constraints and maintain stability of the code. If the user-selected time step for the SDM implementation is too large, collisional breakup may quickly become a runaway process with superdroplet multiplicities increasingly rapidly and unphysically, leading to numerical overflow. As an example, suppose a droplet of multiplicity 10^2 should undergo 6 collisional breakups ($\gamma_\alpha = 5$) into 5 fragments each time ($N_{f,\alpha} = 5$): then $\gamma_{jk} = 3906$ and its new multiplicity is $15,525 = \mathcal{O}(10^4)$. Successive collisional breakups between droplets whose multiplicities have grown so rapidly would then lead to exponentially booming multiplicities, and could quickly exceed the maximum representable quantity for the computing machine (overflow). One solution is to set a maximum allowable multiplicity for any superdroplet, and to reject any collisional breakups that would produce a superdroplet exceeding this multiplicity.

In addition, the process of collisional breakup is physically constrained such that the resulting superdroplet (the "fragment") volume should not exceed the volume of the coalesced transition state, nor should it drop below a realistic size for a liquid water droplet (molecule scale, for instance). These physical constraints can be imposed by setting a minimum and maximum allowable fragment size resulting from breakup.

The first of these constraints can then be imposed during computation of γ_{jk} within the while loop in equation 2.7:

$$\xi_k^{\text{new}} \leq \xi_{\text{max}}, \tag{2.14}$$

where ξ_{max} is a maximum multiplicity set to prevent overflow. The second two constraints are imposed during the sampling of a fragment size:

$$m_{\text{min}} \leq m_{f,\alpha} \leq m_j + m_k, \tag{2.15}$$

where m_{min} is a minimum physically allowed fragment size.

2.6.1 Sampling from empirical fragment size distributions

Sampling a fragment size $m_{f,\alpha}$ requires selecting at random one $m_{f,\alpha}$ according to its weight in the overall fragment size distribution. For fragmentation functions which are composed of several distinct modes, modes must be additionally mass-weighted according to the mode's associated fragment sizes in order to attain the correct expectation value of fragment mass in the resulting sample. A common way of sampling from simple distributions, such as a normal or lognormal, is to invert the CDF (cumulative distribution function). We will demonstrate how to extend this procedure to an empirical fragmentation function comprised of several modes, which lacks a closed form CDF. For instance, the commonly-used fragmentation function of Low and List partitions the fragmentation physics into three categories, corresponding to filament, sheet, and disk breakup. Each category is then comprised of 2–3 modes, corresponding to smaller satellite fragments and larger remnants of the parent droplets. Similarly, Straub et al. distinguish four categories of fragments, with the fragment size distribution within each category following a lognormal or normal distribution. (Here, we follow the notation of Straub et al.)

Suppose the unnormalized fragmentation function $P_f(D)$ in droplet diameter D is described as a sum of k modes:

$$P_f(D) = \sum_{r=1}^{k} N_r p_r(D), \tag{2.16}$$

where N_r is the expected number of fragments from mode r, and $p_r(D)$ is the normalized fragment size distribution for mode r. We begin by re-weighting the size distribution by the mass contained within each mode, M_r:

$$M_r = \int_0^\infty \frac{\pi}{6} D^3 p_r(D) dD, \tag{2.17}$$

and then normalizing the distribution such that its integral is unity:

$$P_{f,n}(D) = \frac{\sum_{r=1}^{k} N_r M_r p_r(D)}{\sum_{r=1}^{k} N_r M_r}. \tag{2.18}$$

These initial transformations yield a normalized fragment size distribution which contains all modes of the fragmentation function weighted according to mass. $P_{f,n}(D)$ corresponds to the probability (as a fraction) of retrieving a sample of mass D given the colliding droplet parameters.

Next, to sample a single fragment size D_f, we first use the random number ϕ''_α to determine which mode of the overall distribution is sampled. This equates to finding s such that

$$\frac{\sum_{r=1}^{s-1} N_r M_r}{\sum_{r=1}^{k} N_r M_r} \leq \phi''_\alpha < \frac{\sum_{r=1}^{s} N_r M_r}{\sum_{r=1}^{k} N_r M_r}. \tag{2.19}$$

The resulting fragment will represent only a single mode of the overall fragment size distribution. However, selecting the mode according to its mass-weighted probability conserves the expected mass distribution of the fragmentation function. The average of several such steps is more likely to sample from each mode, therefore it is crucial that a small enough time step is chosen to allow convergence of this stochastic selection across collisions.

Next, the fragment size is chosen by sampling at random from the CDF of the mode $p_s(D)$, which is assumed to be approximable by a closed form equation (as in the case of a Gaussian or lognormal distribution). This second step of sampling can be accomplished by selecting a new random number, reusing the random number from a different colliding droplet pair, or simply rescaling ϕ''_α as

$$\tilde{\phi}''_\alpha = \frac{\phi''_\alpha - \sum_{r=1}^{s-1} N_r M_r}{\sum_{r=1}^{s} N_r M_r - \sum_{r=1}^{s-1} N_r M_r}. \tag{2.20}$$

The fragment size D_f is then selected such that

$$\tilde{\phi}''_\alpha = p_s(D_f). \tag{2.21}$$

There are several methods of sampling a fragment size from distributions composed of several modes. The presented implementation is used in generating results in this work, and is included as one such example.

Author Contributions

EdJ led the code development, generation of results, interpretation, and writing. BM contributed to code development and the underlying methodology. OB contributed to algorithm and code review, validation against analytic solution and to porting the code for the GPU backend of PySDM. AJ contributed to underlying methodology and interpretation of results. SA contributed to the code development, interpretation, and leads maintenance of the PySDM codebase.

Acknowledgements

We thank Tapio Schneider and Shin-ichiro Shima for feedback, insights, and discussion. The study benefited from peer-review comments provided by Axel Seifert, Christoph Siewert and two anonymous reviewers. Additional thanks go to Piotr Bartman for ongoing support of PySDM. E. de Jong was supported by a Department of Energy Computational Sciences Graduate Fellowship. S. Arabas and O. Bulenok acknowledge support from the Polish National Science Centre (grant no. 2020/39/D/ST10/01220). This research was additionally supported by Eric and Wendy Schmidt (by recommendation of Schmidt Futures) and the Heising-Simons Foundation.

References

Andrejczuk, M., Grabowski, W. W., Reisner, J., & Gadian, A. (2010). Cloud-aerosol interactions for boundary layer stratocumulus in the Lagrangian Cloud Model. *Journal of Geophysical Research: Atmospheres*, *115*(22). https://doi.org/10.1029/2010JD014248

Andrejczuk, M., Reisner, J. M., Henson, B., Dubey, M. K., & Jeffery, C. A. (2008). The potential impacts of pollution on a nondrizzling stratus deck: Does aerosol number matter more than type? *Journal of Geophysical Research: Atmospheres*, *113*(D19). https://doi.org/10.1029/2007JD009445

Arabas, S., Jaruga, A., Pawlowska, H., & Grabowski, W. (2015). Libcloudph++ 1.0: A single-moment bulk, double-moment bulk, and particle-based warm-rain microphysics library in c++. *Geosci. Model Dev.*, *8*. https://doi.org/10.5194/gmd-8-1677-2015

Arabas, S., Bartman, P., Jong, E. d., Singer, C., Olesik, M. A., Mackay, B., Bulenok, O., Azimi, S., Górski, K., Jaruga, A., Piasecki, B., & Badger, C. (2022, August). Atmos-cloud-sim-uj/PySDM: PySDM v2.12. Retrieved October 21, 2022, from https://zenodo.org/record/7037182

Arenberg, D. (1939). Turbulence As The Major Factor in the Growth of Cloud Drops. *Bulletin of the American Meteorological Society*, *20*(10), 444–448. https://doi.org/10.1175/1520-0477-20.10.444

Barros, A. P., Prat, O. P., Shrestha, P., Testik, F. Y., & Bliven, L. F. (2008). Revisiting Low and List (1982): Evaluation of Raindrop Collision Parameterizations Using Laboratory Observations and Modeling. *Journal of the Atmospheric Sciences*, *65*(9), 2983–2993. https://doi.org/10.1175/2008JAS2630.1

Bartman, P., Banaśkiewicz, J., Drenda, S., Manna, M., Olesik, M. A., Rozwoda, P., Sadowski, M., & Arabas, S. (2022). PyMPDATA v1: Numba-accelerated implementation of MPDATA with examples in Python, Julia and Matlab. *Journal of Open Source Software*, *7*(77), 3896. https://doi.org/10.21105/joss.03896

Bartman, P., Bulenok, O., Górski, K., Jaruga, A., Łazarski, G., Olesik, M. A., Piasecki, B., Singer, C. E., Talar, A., & Arabas, S. (2022). PySDM v1: Particle-based cloud modeling package for warm-rain microphysics and aqueous chemistry. *Journal of Open Source Software*, *7*(72), 3219. https://doi.org/10.21105/joss.03219

Beard, K. V., & Ochs, H. T. (1995). Collisions between Small Precipitation Drops. Part II: Formulas for Coalescence, Temporary Coalescence, and Satellites. *Journal of the Atmospheric Sciences*, *52*(22), 3977–3996. https://doi.org/10.1175/1520-0469(1995)052<3977:CBSPDP>2.0.CO;2

Berry, E. X. (1967). Cloud Droplet Growth by Collection. *Journal of Atmospheric Sciences*, *24*(6), 688–701. https://doi.org/10.1175/1520-0469(1967)024<0688:CDGBC>2.0.CO;2

Bieli, M., Dunbar, O. R. A., de Jong, E. K., Jaruga, A., Schneider, T., & Bischoff, T. (2022). An Efficient Bayesian Approach to Learning Droplet Collision Kernels: Proof of Concept Using "Cloudy," a New n-Moment Bulk Microphysics Scheme. *Journal of Advances in Modeling Earth Systems*, *14*(8), e2022MS002994. https://doi.org/10.1029/2022MS002994

Blatz, P., & Tobolsky, A. (1945). Note on the kinetics of systems manifesting simultaneous polymerization-depolymerization phenomena. *J. Phys. Chem.*, *49*. https://doi.org/10.1021/j150440a004

Bringi, V., Seifert, A., Wu, W., Thurai, M., Huang, G.-J., & Siewert, C. (2020). Hurricane Dorian Outer Rain Band Observations and 1D Particle Model Simulations: A Case Study [Number: 8 Publisher: Multidisciplinary Digital Publishing Institute]. *Atmosphere*, *11*(8), 879. https://doi.org/10.3390/atmos11080879

Chandrakar, K. K., Grabowski, W. W., Morrison, H., & Bryan, G. H. (2021). Impact of Entrainment Mixing and Turbulent Fluctuations on Droplet Size Distributions in a Cumulus Cloud: An Investigation Using Lagrangian Microphysics with a Subgrid-Scale Model. *Journal of the Atmospheric Sciences*, *78*(9), 2983–3005. https://doi.org/10.1175/JAS-D-20-0281.1

de Jong, E. K., Singer, C. E., Azimi, S., Bartman, P., Bulenok, O., Derlatka, K., Dula, I., Jaruga, A., Mackay, J. B., Ward, R. X., & Arabas, S. (2023). New developments in pysdm and pysdm-examples v2: Collisional breakup, immersion freezing, dry aerosol initialization, and adaptive time-stepping. *Journal of Open Source Software*, *8*(84), 4968. https://doi.org/10.21105/joss.04968

Dziekan, P., Waruszewski, M., & Pawlowska, H. (2019). University of Warsaw Lagrangian Cloud Model (UWLCM) 1.0: A modern large-eddy simulation tool for warm cloud modeling with Lagrangian microphysics. *Geoscientific Model Development*, *12*(6), 2587–2606. https://doi.org/10.5194/gmd-12-2587-2019

Feingold, G., Cotton, W. R., Kreidenweis, S. M., & Davis, J. T. (1999). The Impact of Giant Cloud Condensation Nuclei on Drizzle Formation in Stratocumulus: Implications for Cloud Radiative Properties. *Journal of the Atmospheric Sciences*, *56*(24), 4100–4117. https://doi.org/10.1175/1520-0469(1999)056<4100:TIOGCC>2.0.CO;2

Grabowski, W. W. (2020). Comparison of Eulerian Bin and Lagrangian Particle-Based Schemes in Simulations of Pi Chamber Dynamics and Microphysics. *Journal of the Atmospheric Sciences*, *77*(3), 1151–1165. https://doi.org/10.1175/JAS-D-19-0216.1

Grabowski, W. W., Morrison, H., Shima, S.-I., Abade, G. C., Dziekan, P., & Pawlowska, H. (2019). Modeling of Cloud Microphysics: Can We Do Better? *Bulletin of the American Meteorological Society*, *100*(4), 655–672. https://doi.org/10.1175/BAMS-D-18-0005.1

Gunn, R., & Kinzer, G. D. (1949). The Terminal Velocity of Fall for Water Droplets in Stagnant Air. *Journal of the Atmospheric Sciences*, *6*(4), 243–248. https://doi.org/10.1175/1520-0469(1949)006<0243:TTVOFF>2.0.CO;2

Hallett, J., & Mossop, S. C. (1974). Production of secondary ice particles during the riming process. *Nature*, *249*(5452), 26–28. https://doi.org/10.1038/249026a0

Hansen, K. (2018). Abundance distributions; large scale features. In *Statistical physics of nanoparticles in the gas phase*. https://doi.org/10.1007/978-3-319-90062-9_8

Harris-Hobbs, R. L., & Cooper, W. A. (1987). Field Evidence Supporting Quantitative Predictions of Secondary Ice Production Rates. *Journal of the Atmospheric Sciences*, *44*(7), 1071–1082. https://doi.org/10.1175/1520-0469(1987)044<1071:FESQPO>2.0.CO;2

Hoffmann, F. (2017). On the limits of Köhler activation theory: How do collision and coalescence affect the activation of aerosols? *Atmospheric Chemistry and Physics*, *17*(13), 8343–8356. https://doi.org/https://doi.org/10.5194/acp-17-8343-2017

James, R. L., Phillips, V. T. J., & Connolly, P. J. (2021). Secondary ice production during the break-up of freezing water drops on impact with ice particles. *Atmospheric Chemistry and Physics*, *21*(24), 18519–18530. https://doi.org/10.5194/acp-21-18519-2021

Jensen, E., & Pfister, L. (2004). Transport and freeze-drying in the tropical tropopause layer. *Journal of Geophysical Research: Atmospheres*, *109*(D2). https://doi.org/10.1029/2003JD004022

Jokulsdottir, T., & Archer, D. (2016). A stochastic, Lagrangian model of sinking biogenic aggregates in the ocean (SLAMS 1.0): Model formulation, validation and sensitivity. *Geoscientific Model Development*, *9*(4), 1455–1476. https://doi.org/10.5194/gmd-9-1455-2016

Kamra, A. K., Bhalwankar, R. V., & Sathe, A. B. (1991). Spontaneous breakup of charged and uncharged water drops freely suspended in a wind tunnel. *Journal of Geophysical Research: Atmospheres*, *96*(D9), 17159–17168. https://doi.org/10.1029/91JD01475

Kotalczyk, G., Devi, J., & Kruis, F. E. (2017). A time-driven constant-number Monte Carlo method for the GPU-simulation of particle breakage based on weighted simulation particles. *Powder Technology*, *317*, 417–429. https://doi.org/10.1016/j.powtec.2017.05.002

Lee, K., & Matsoukas, T. (2000). Simultaneous coagulation and break-up using constant-N Monte Carlo. *Powder Technology*, *110*(1), 82–89. https://doi.org/10.1016/S0032-5910(99)00270-3

Low, T. B., & List, R. (1982). Collision, Coalescence and Breakup of Raindrops. Part I: Experimentally Established Coalescence Efficiencies and Fragment Size Distributions in Breakup. *Journal of the Atmospheric Sciences*, *39*(7), 1591–1606. https://doi.org/10.1175/1520-0469(1982)039<1591:CCABOR>2.0.CO;2

McFarquhar, G. M. (2004). A New Representation of Collision-Induced Breakup of Raindrops and Its Implications for the Shapes of Raindrop Size Distributions. *Journal of the Atmospheric Sciences*, *61*(7), 777–794. https://doi.org/10.1175/1520-0469(2004)061<0777:ANROCB>2.0.CO;2

Morrison, H., Kumjian, M. R., Martinkus, C. P., Prat, O. P., & van Lier-Walqui, M. (2019). A General N-Moment Normalization Method for Deriving Raindrop Size Distribution Scaling Relationships. *Journal of Applied Meteorology and Climatology*, *58*(2), 247–267. https://doi.org/10.1175/JAMC-D-18-0060.1

Morrison, H., van Lier-Walqui, M. v., Fridlind, A. M., Grabowski, W. W., Harrington, J. Y., Hoose, C., Korolev, A., Kumjian, M. R., Milbrandt, J. A., Pawlowska, H., Posselt, D. J., Prat, O. P., Reimel, K. J., Shima, S.-I., Diedenhoven, B. v., & Xue, L. (2020). Confronting the Challenge of Modeling Cloud and Precipitation Microphysics. *Journal of Advances in Modeling Earth Systems*, *12*(8). https://doi.org/10.1029/2019MS001689 e2019MS001689 2019MS001689.

Paoli, R., Hélie, J., & Poinsot, T. (2004). Contrail formation in aircraft wakes. *Journal of Fluid Mechanics*, *502*, 361–373. https://doi.org/10.1017/S0022112003007808

Phillips, V. T. J., Yano, J.-I., & Khain, A. (2017). Ice Multiplication by Breakup in Ice–Ice Collisions. Part I: Theoretical Formulation. *Journal of the Atmospheric Sciences*, *74*(6), 1705–1719. https://doi.org/10.1175/JAS-D-16-0224.1

Riechelmann, T., Noh, Y., & Raasch, S. (2012). A new method for large-eddy simulations of clouds with Lagrangian droplets including the effects of turbulent collision. *New Journal of Physics*, *14*(6), 065008. https://doi.org/10.1088/1367-2630/14/6/065008

Schlottke, J., Straub, W., Beheng, K. D., Gomaa, H., & Weigand, B. (2010). Numerical Investigation of Collision-Induced Breakup of Raindrops. Part I: Methodology and Dependencies on Collision Energy and Eccentricity. *Journal of the Atmospheric Sciences*, *67*(3), 557–575. https://doi.org/10.1175/2009JAS3174.1

Seifert, A., Khain, A., Blahak, U., & Beheng, K. D. (2005). Possible Effects of Collisional Breakup on Mixed-Phase Deep Convection Simulated by a Spectral (Bin) Cloud Model. *Journal of the Atmospheric Sciences*, *62*(6), 1917–1931. https://doi.org/10.1175/JAS3432.1

Seifert, A., & Rasp, S. (2020). Potential and Limitations of Machine Learning for Modeling Warm-Rain Cloud Microphysical Processes. *Journal of Advances in Modeling Earth Systems*, *12*(12), e2020MS002301. https://doi.org/https://doi.org/10.1029/2020MS002301

Shima, S.-I., Kusano, K., Kawano, A., Sugiyama, T., & Kawahara, S. (2009). The super-droplet method for the numerical simulation of clouds and precipitation: A particle-based and probabilistic microphysics model coupled with a non-hydrostatic model. *Quarterly Journal of the Royal Meteorological Society*, *135*(642), 1307–1320. https://doi.org/10.1002/qj.441

Shima, S.-I., Sato, Y., Hashimoto, A., & Misumi, R. (2020). Predicting the morphology of ice particles in deep convection using the super-droplet method: Development and evaluation of SCALE-SDM 0.2.5-2.2.0, -2.2.1, and -2.2.2. *Geoscientific Model Development*, *13*(9), 4107–4157. https://doi.org/10.5194/gmd-13-4107-2020

Shipway, B. J., & Hill, A. A. (2012). Diagnosis of systematic differences between multiple parametrizations of warm rain microphysics using a kinematic framework. *Quarterly Journal of the Royal Meteorological Society*, *138*(669), 2196–2211. https://doi.org/10.1002/qj.1913

Shirgaonkar, A., & Lele, S. (2006, January). Large Eddy Simulation of Early Stage Contrails: Effect of Atmospheric Properties. In *44th AIAA Aerospace Sciences Meeting and Exhibit*. American Institute of Aeronautics; Astronautics. Retrieved October 21, 2022, from https://arc.aiaa.org/doi/10.2514/6.2006-1414

Sölch, I., & Kärcher, B. (2010). A large-eddy model for cirrus clouds with explicit aerosol and ice microphysics and Lagrangian ice particle tracking. *Quarterly Journal of the Royal Meteorological Society*, *136*(653), 2074–2093. https://doi.org/10.1002/qj.689

Srivastava, R. (1982). A simple model of particle coalescence and breakup. *J. Atmos. Sci.*, *39*. https://doi.org/10.1175/1520-0469(1982)039<1317:ASMOPC>2.0.CO;2

Straub, W., Beheng, K. D., Seifert, A., Schlottke, J., & Weigand, B. (2010). Numerical Investigation of Collision-Induced Breakup of Raindrops. Part II: Parameterizations of Coalescence Efficiencies and Fragment Size Distributions. *Journal of the Atmospheric Sciences*, *67*(3), 576–588. https://doi.org/10.1175/2009JAS3175.1

Testik, F. Y., & Rahman, M. K. (2017). First in situ observations of binary raindrop collisions. *Geophysical Research Letters*, *44*(2), 1175–1181. https://doi.org/10.1002/2017GL072516

Wood, R., Irons, S., & Jonas, P. R. (2002). How Important Is the Spectral Ripening Effect in Stratiform Boundary Layer Clouds? Studies Using Simple Trajectory Analysis. *Journal of the Atmospheric Sciences*, *59*(18), 2681–2693. https://doi.org/10.1175/1520-0469(2002)059<2681:HIITSR>2.0.CO;2

Yin, Y., Levin, Z., Reisin, T. G., & Tzivion, S. (2000). The effects of giant cloud condensation nuclei on the development of precipitation in convective clouds — a numerical study. *Atmospheric Research*, *53*(1), 91–116. https://doi.org/10.1016/S0169-8095(99)00046-0

Zhao, X., & Liu, X. (2022). Primary and secondary ice production: Interactions and their relative importance. *Atmospheric Chemistry and Physics*, *22*(4), 2585–2600. https://doi.org/10.5194/acp-22-2585-2022

Chapter 3

SPANNING THE GAP FROM BULK TO BIN: A NOVEL SPECTRAL MICROPHYSICS METHOD

de Jong, E. K., Bischoff, T., Nadim, A., & Schneider, T. (2022). Spanning the gap from bulk to bin: A novel spectral microphysics method. *Journal of Advances in Modeling Earth Systems*, *14*(11), e2022MS003186. https://doi.org/10.1029/2022MS003186

Abstract

Microphysics methods for climate models and numerical weather prediction typically track one, two, or three moments of a droplet size distribution for various categories of liquid, ice, and aerosol. Such methods rely on conversion parameters between these categories, which introduces uncertainty into predictions. While higher-fidelity options such as bin and Lagrangian schemes exist, they require too many degrees of freedom for climate modeling applications and introduce numerical challenges. Here we introduce a flexible spectral microphysics method based on collocation of basis functions. This method generalizes to a linear bulk scheme when using few basis functions and to a smoothed bin scheme with more degrees of freedom. Tested in an idealized box setting, the method improves spectral accuracy for droplet collision-coalescence and may eliminate the need for precipitation autoconversion rates required by bulk methods; furthermore, it generalizes well to multimodal distributions with less complexity than a bin method. The potential to extend this collocation representation to multiple hydrometeor classes suggests a path forward to unify liquid, ice, and aerosol microphysics in a single, flexible, computational framework for climate modeling.

Plain Language Summary

Clouds and aerosols affect global warming by reflecting and absorbing radiation and by storing and transporting water. Climate models need a way to efficiently track the size and number of cloud droplets, ice, and aerosols in order to accurately predict the impact that these “microphysical” particles have on climate. Existing methods of microphysics rely on many uncertain parameters and are either too complicated or too simple to take advantage of today’s computational resources. We propose a

new way to represent cloud droplets that can both reduce uncertainties and make use of increased computing power.

3.1 Introduction

Droplets, aerosols, and ice particles, collectively a subset of atmospheric microphysical particles, affect planetary-scale climate, yet the processes that govern their behavior occur at the microscale. This extreme range of scales, from droplets to clouds to large-scale atmospheric dynamics, makes it challenging to computationally represent microphysics in atmosphere simulations. There are simply too many particles to represent directly, yet the microphysics processes involved are nonlinear and do not lend themselves easily to simplifications. Instead, microphysics schemes in climate and numerical weather models predict the particle size distribution (PSD) present at various locations in the atmosphere: the PSD and number concentration determine the macroscopic behavior of the system, such as cloud albedo or precipitation rates. Historically, methods to represent the PSD developed along two trajectories: bulk methods, which predict aggregate properties of the droplet population, and spectral methods, which explicitly track the PSD. Both of these representations make assumptions about the droplet distribution and the microphysical process rates, with spectral methods being the more flexible of the two options. Unfortunately, these parameterizations and assumptions contribute a major yet difficult-to-quantify source of uncertainty in climate predictions (Arakawa, 2004; Intergovernmental Panel on Climate Change, 2014; Khain et al., 2015; Morrison et al., 2020; Randall et al., 2003).

Bulk schemes, originating with Kessler, 1969, explicitly track one or more prognostic moments of the PSD and therefore are compact representations suitable for global climate applications. Indeed, many diagnostic cloud quantities, such as the radar reflectivity or liquid mass fraction are directly related to these moments. On the other hand, many process rates such as collisional coalescence depend on the full PSD, which is not uniquely defined by the prognostic moments. Therefore, by abstracting a droplet population to one, two, or three variables, bulk methods make two fundamental simplifications. First, many single-droplet processes such as sedimentation or aerosol activation require additional parameterizations or assumptions to approximate how the processes impact the prognostic moments. Second, because many such process rates depend on higher-order moments which are not explicitly tracked, moment-based methods require a closure to relate these higher order moments back to the prognostic variables. Frequently this closure is accomplished by relating the prognostic moments back to an underlying assumed size distribution such as a gamma or exponential Milbrandt and Yau, 2005; Morrison and Grabowski, 2008; Seifert and Beheng, 2006, which corresponds well to data in many empirical settings. However, in the case of a multimodal distribution, (for instance, when both small cloud droplets and larger rain droplets are present) this closure assumption introduces significant structural uncertainty into the microphysics scheme. There is no physical reason *a priori* to restrict a droplet population to maintaining a particular size distribution as they coalesce, break up, grow, sediment, and change phases. Unfortunately, inverting a multimodal distribution analytically is frequently ill-posed (Morrison et al., 2019). Most traditional bulk methods avoid the issue by representing several categories of hydrometeors (rain, cloud droplets, and several categories of aerosols) through separate prognostic moments, assuming a simple unimodal distribution for each of these categories. However, these categories of condensed water, while intuitive, are artificial: in reality, liquid hydrometeors are distributed across a continuous spectrum, from small chemically-active aerosol particles, to large liquid cloud droplets, to droplets which are large enough to fall as rain. Conversion between these categories adds further complexity and uncertainty to the model.

On the other hand, spectral or "bin" microphysics schemes directly evolve the PSD in time through discrete bins, or particle size ranges (e.g. Berry, 1967; Berry and Reinhardt, 1974; Tzivion (Tzitzvashvili) et al., 1987; Young, 1974). Bin methods have made a great impact in understanding aerosol-cloud interactions (e.g. Fan et al., 2016; Khain et al., 2015; Morrison and Grabowski, 2007), but at a higher computational cost that currently makes them infeasible for climate simulations. For example, Gettelman et al., 2021 ran a general circulation model (GCM) with bin microphysics, incurring a factor of five cost penalty over a bulk scheme. Furthermore, while bin methods avoid the closure assumptions of bulk schemes, they suffer from similar parameterization challenges, numerical diffusion (Morrison et al., 2019), as well as from sensitivity to the bin discretization (Ghan et al., 2011). The purpose of the method presented here is to target the middle ground of complexity between traditional bulk and bin methods using more sophisticated numerical techniques.

To meet the needs of future climate and weather models, a microphysics scheme should maintain enough flexibility to function with a wide range of degrees of freedom and minimal structural uncertainty in the PSD representation. While bin-scheme complexity may be unattainable for GCMs in the near future, we still need a microphysics method that can maintain spectral details without the closure assumptions and conversion parameterizations required by moment-based bulk methods. Some recent efforts in microphysics modeling have focused on relaxing assumptions about the size distribution and process rates to reduce these structural uncertainties. One option, Lagrangian microphysics, directly tracks tracer particles known as superdroplets (Andrejczuk et al., 2008, 2010; Riechelmann et al., 2012; Shima et al., 2009), but it is far too computationally expensive for global or even regional models. A different moment-based method, the BOSS scheme proposed by Morrison et al., 2019 leaves all process rates and closures as generalized power series whose parameters are learned from data. Bieli et al., 2022 present a more efficient way to learn these parameters within a similar bulk microphysics framework that still relies on closures. More complex yet, Rodríguez Genó and Alfonso, 2022 tackle the challenge of inverting multimodal distribution closures using a machine-learning based method, which could avoid the necessity for cloud-rain conversion rate parameterizations. Another novel approach of combining bulk and bin microphysics to generate arbitrary moments for potentially multimodal distributions (Igel, 2019) illustrates the need to relax finite-size threshold assumptions for cloud-to-rain conversion (Igel et al., 2022). However, these bulk methods cannot function in a wide range of computational degrees of freedom, nor do they provide complete spectral details about the PSD that might alleviate uncertainties about conversion between hydrometeor types. One solution is to think beyond the classical bulk versus bin representations of the PSD, leveraging numerical techniques developed for fluid mechanics.

In this study, we present and test a novel way to span the gap in complexity between bin and bulk microphysics methods by applying the collocation method with basis functions (BFs) to represent the particle size distribution. (For simplicity, it will be referred to going forward as the BF method.) Finite element methods such as collocation have been historically overlooked for microphysics applications, with the exception of Gelbard and Seinfeld, 1978's demonstration using collocation of quartic or cubic polynomials, which was never widely adopted in favor of contemporaneous bin methods. More recent results from the applied math community suggest that combining collocation with radial basis functions, rather than polynomials, is a promising numerical technique for advection problems (Franke & Schaback, 1998; Zhang et al., 2000). This work extends the basis function collocation technique to the integro-differential equations encountered in microphysics. Beyond retaining spectral details of the PSD, the BF method has appealing extremes of complexity: when using few basis functions, the method is effectively a linear closure, as in the context of bulk schemes; at moderate or high resolutions, it converges toward a smoothed bin scheme (replicating a bin scheme exactly if constant piecewise BFs and appropriate numerics are used). Therefore collocation of basis functions promises greater flexibility than either bulk or bin methods alone, while retaining desirable aspects such as low-to-moderate complexity and spectral predictions. This paper describes the method and presents results of applying the method to droplet collision and coalescence, benchmarked against commonly used bulk, bin, and Lagrangian frameworks. We additionally address some limitations posed by the method that are specific to the context of tracking a PSD, such as mass non-conservation and a finite size range. Overall, the BF method improves spectral PSD predictions in a box model as well as simple precipitation predictions, measured as a size exceedance, compared to a three-moment bulk method, and with fewer degrees of freedom than a bin method. Furthermore, the run-time complexity of the method scales quadratically with the number of degrees of freedom, making it just as efficient as a bin method.

The remainder of this paper is organized as follows: section 2 describes the method of collocation of basis functions to approximately solve the population balance equation for collision-coalescence in microphysics, and section 3 describes a set of microphysics box model case studies. Section 4 compares the accuracy of these case studies solved using basis functions, bulk, and bin methods against a high-fidelity Lagrangian reference solution, and discusses the computational complexity of these methods. Finally, section 5 concludes the paper and suggests potential improvements and applications.

3.2 Method Description

3.2.1 Key Equations

The governing equations for microphysics describe a population balance for the droplet size distribution. The governing equation for collision-coalescence, also called the Smoluchowski or Stochastic Collection Equation (SCE), is given by

$$\begin{aligned}\partial_t n(x,t) = \frac{1}{2}\int_0^x & n(x-y,t)n(y,t)K(x-y,y)E_c(x-y,y)dy \\ & -n(x)\int_0^\infty n(y,t)K(x,y)E_c(x,y)dy,\end{aligned} \tag{3.1}$$

where $n(x,t)$ represents the number density of particles of mass x at time t, $K(x,y)$ is the collision rate of particles of masses x and y, and $E_c(x,y)$ is the coalescence efficiency of said collision. The first integral represents production of droplets of size x from two smaller droplets, and the second integral represents loss of droplets of size x due to coalescence with other droplets.

Other microphysical processes such as condensation, evaporation, sedimentation, and aerosol activation also affect the PSD. To demonstrate the proposed BF method for microphysics, we initially focus on only the coalescence process as in equation (3.1). The SCE is notoriously difficult to solve numerically, as it is an integro-partial differential equation and frequently involves rapid acceleration of particle growth, yet this mechanism is crucial to determining the onset of rain and drizzle (Stephens et al., 2010). Later, we will also consider two non-collisional processes of sedimentation and injection of new particles. For the purposes of this study, sedimentation is defined as removal of all particles above a size threshold x_{max}, which can prevent unphysically rapid acceleration of collisions. Sedimentation is enforced by limiting the upper bound of each integral to x_{max}, effectively truncating the PSD to have a value of $n(x > x_{\text{max}}, t) = 0$. We can alternatively prevent particles larger than the maximum size x_{max} from forming by rejecting those collisions in a mass conserving manner. The appropriate upper bound for the second integral in this case is $x_{\text{max}} - x$ (Filbet & Laurençot, 2004). When such collisions are not rejected and particles exit the system, we introduce new droplets to the system, to mimic the entrainment or activation of new small particles. The rate of particle injection $P_{\text{inj}}(x,t)$ is given by

$$P_{\text{inj}}(x,t) = \dot{P} I(x) \tag{3.2}$$

where $I(x)$ represents a normalized size distribution of the injected droplets, which might be smaller than the average droplet in the system, and $\dot{P}$ is the rate of particle injection. This combination of droplet dynamics represents a system in which particles continuously enter the system, grow through coalescence, and exit by sedimentation once they reach a critical size.

3.2.2 Collocation of Basis Functions with Positivity Constraint

In our proposed method, based on the work of Zhang et al., 2000, the PSD is approximated by a weighted sum of n_{BF} basis functions:

$$n(x,t) \approx \tilde{n}(x,t) = \sum_{k=1}^{n_{\text{BF}}} c_k(t)\phi(x;\theta_k) = \mathbf{c}(t) \cdot \boldsymbol{\phi}(x). \tag{3.3}$$

We denote the approximate solution $\tilde{n}(x,t)$, the collocation weights $c_k(t)$, and the basis functions $\phi(x|\theta_k)$ where ϕ is the functional form and θ_k are the parameters of the k-th BF (for instance, mean and variance of a Gaussian). In the collocation method, one such parameter is the center or mean of the basis function, $\mu_k \in \theta_k$, known as the collocation points. In the context of microphysics, these collocation points refer to particle masses, which locate the mode of each basis function. In equation (3.3), we have also compactly rewritten the BFs and weights in vector form as $\boldsymbol{\phi}(x) = \big(\phi(x|\theta_1), \phi(x|\theta_2), \ldots, \phi(x|\theta_{n_{\mathrm{BF}}})\big)$ and $\mathbf{c}(t) = \big(c_1(x), c_2(x), \ldots, c_{n_{\mathrm{BF}}}(x)\big)$.

Since the basis functions have a fixed shape over the droplet size range, evolving the approximate PSD reduces to solving for $\mathbf{c}(t)$ in time as a system of ordinary differential equations. Because liquid water is a conserved quantity in the absence of evaporation/condensation, we consider the evolution of the local mass density $m(x,t) = x\,n(x,t)$ rather than the local number density. Thus although we use basis functions to approximate the number density, the equations are evolved in time based on local mass density, as in a one-moment bulk method or a standard flux-method bin scheme.

Denote the vector of approximate mass density at the collocation points μ_k to be $\tilde{\mathbf{m}}(t) = \big(\mu_1\tilde{n}(\mu_1,t), \ldots, \mu_p\tilde{n}(\mu_p,t)\big)$. At each timestep, recovering the weights from the interpolated collocation points requires solving for $\mathbf{c}(t)$ in the linear system

$$\tilde{\mathbf{m}}(t) = \mathbf{\Phi} \cdot \mathbf{c}(t), \tag{3.4}$$

where $\mathbf{\Phi}$ is a $n_{\mathrm{BF}} \times n_{\mathrm{BF}}$ matrix, with elements $\Phi_{jk} = \mu_j\phi_k(\mu_j)$ representing the mass density of the k-th basis function evaluated at the jth collocation point. For a linearly independent set of basis functions, this system is well-posed and guarantees a unique solution. However, it may be ill-conditioned, particularly when the choice of basis function has global rather than compact support (Zhang et al., 2000).

The approximate solution is initialized by projecting the initial mass distribution onto the basis space using the collocation points, as in Equation 3.4. This system is solved via a constrained optimization problem:

$$\min_{\mathbf{c}(t)} \quad \|\mathbf{\Phi}\cdot\mathbf{c}(t) - \tilde{\mathbf{m}}(t)\|^2 \qquad \text{s.t.} \quad \mathbf{c}(t) \geq 0\,. \tag{3.5}$$

The positivity constraint mathematically enforces the fact that the PSD should be nonnegative at all points. Equation (3.5) is formulated as a quadratic optimization, and therefore can be solved efficiently via least squares.

This projection could additionally incorporate a mass conservation constraint, both initially and at every future time step, but at higher cost than solving the linear system in equation 3.4. Additionally, since the exact solution to the equation does not necessarily exist as a projection of the basis functions, the mass and positivity constraints in the optimizer can lead to unphysical solutions as the approximate PSD evolves in time. While relaxing this constraint might lead to an artificial reduction or increase in mass throughout the simulation time, it allows a more efficient nonnegative least-squares solution. In developing this method, we observed that evolving the linear system in mass density with a positivity constraint, rather than using number density directly, led to more physical and realistic PSDs compared to including a mass-conserving constraint at all times.

3.2.2.1 Interpretability and design choices

The method described above generalizes to solve many categories of differential equation, but selecting the basis functions and parameters θ_k requires care in order to preserve physical properties of a droplet distribution, such as physically realistic sizes, finite positive number and mass, and convergence with increasing complexity (for further discussion, refer to Appendix A). To model a droplet PSD, we choose to let the basis functions themselves be distributions, in contrast to the cubic splines employed by Gelbard and Seinfeld, 1978 or spectral element methods. If we choose Gaussian or lognormal BF's collocated on a grid of droplet sizes, each BF effectively represents a droplet size mode. This feature provides a useful analogy to aerosol size modes, or cloud versus rain droplet distributions, much as a typical bin scheme will distinguish between aerosol, cloud, and rain size bins, or how a moment scheme will have a separate set of moments for cloud and rain water. In fact, this representation is a generalization of bin schemes, which can be considered piecewise constant basis functions: $\phi_k(x) = 1, x \in \{x_k, x_{k+1}\}$ (see Figure 3.1). When used with only a few basis functions, the BF representation can similarly be thought of as approximating a linear closure, as in the method of moments (MOM), where the prognostic variable is the first moment calculated over sub-intervals of the particle size range.

Additional design choices include selecting the collocation points and additional hyperparameters of the BFs, such as the variance for lognormal or Gaussian distributions. An in-depth description and justification of the BF setup used in following sections can be found in Appendix A. Notably, we introduce a compactly-supported BF that approximates a lognormal distribution (CSLBF1: equation (3.9)), use

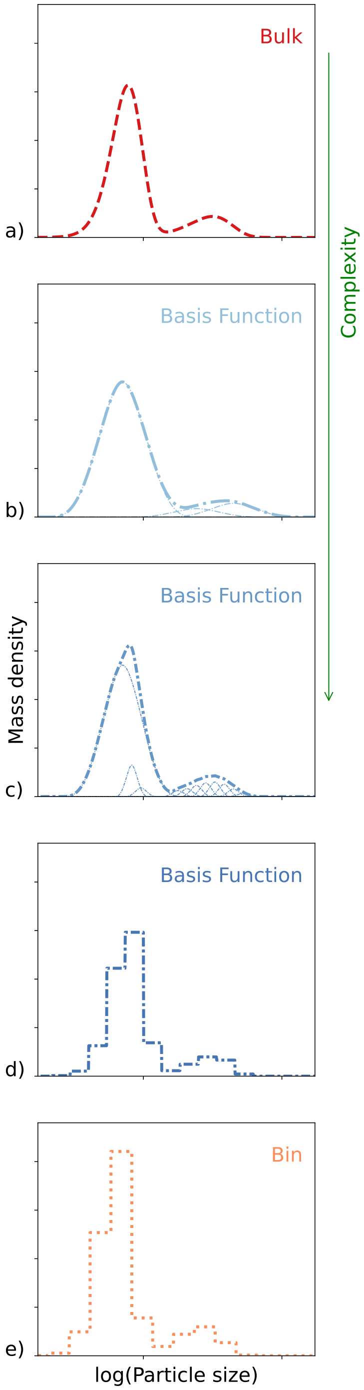


Figure 3.1: Illustration of the way that the collocation of basis functions can span the range in complexity from bulk to bin microphysics. The PSD for a two-mode gamma mixture of particles, corresponding, for instance, to a cloud and rain mode, is plotted as it would be represented in a: (a) 3-moment bulk scheme with gamma closure (one set of moments for each mode); (b) 4 lognormal basis functions; (c) 16 lognormal basis functions; (d) 16 piecewise-constant basis functions; (e) bin method with 32 bins.

exponentially-spaced collocation points, and set the geometric standard deviation as the distance between adjacent collocation points.

3.2.3 Application to the SCE and microphysical processes

The equations involved in applying the BF method to the SCE are derived in appendix B, with the result summarized by equation (3.6) below:

$$\begin{cases} d_t\tilde{\mathbf{m}}(t) = \mathbf{c}(t)\cdot\mathbf{Q}\cdot\mathbf{c}(t) + \sum_{l=1}^{N_{proc}}\mathbf{P}_l \\ \mathbf{\Phi}\cdot\mathbf{c}(t) = \tilde{\mathbf{m}}, \qquad \text{with} \quad \mathbf{c}(t)\geq 0 \end{cases} . \tag{3.6}$$

In this equation, third-order tensor $\mathbf{Q}$ and vectors $\mathbf{P}_l$ are obtained by taking various inner products of the collision kernel and additional process rates (respectively) with the basis functions. All integrals for this collision-coalescence term can be pre-computed for a fixed set of basis functions, defining these tensors through numerical integration and projection of rate processes onto the basis space. (The required precomputations and scaling of these computations with the number of BFs are described in Appendix B. In summary, the precomputation steps scale at most cubically with the number of BFs, and the computation at each time step scales cubically or quadratically depending on the basis chosen.) The result is a simple set of quadratic coupled ordinary differential equations for the mass density at the collocation points, $\tilde{\mathbf{m}}(t)$, and the BF weight vector $\mathbf{c}(t)$. Examples demonstrated in this work use a fourth-order Runge-Kutta time stepper to solve Equation 3.6.

3.3 Test Cases

As a demonstration of the the accuracy, efficiency, and limitations of the proposed BF method, we compare its spectral and bulk quantity accuracy with bin, bulk, and a reference Lagrangian microphysics scheme, using three different initial conditions and four sets of collisional dynamics in a zero-dimensional box. The functional form and parameters of each initial condition are summarized in table 1, and the collision kernels and injection dynamics are summarized in table 2.

The first two initial conditions (table 1) consider a unimodal PSD, following either an exponential (EXP) or gamma (GAM) distribution in particle volume (particles assumed spherical). EXP considers a mean droplet radius of 15 μm, and GAM considers a smaller mean droplet radius of 8.95 μm. The third initial PSD is a bimodal (BIM) sum of two gamma distributions with mean droplet radii 9.85 μm and 4.57 μm. This initial distribution can be thought of as representing two cloud droplet or aerosol modes, or alternatively a cloud mode and rain droplet mode. A

Label	Form	Parameters
EXP	Exponential $n_0(x) = \frac{N_0}{\theta}\exp(-x/\theta)$	$N_0 = 10/\text{cm}^3$ $\theta = 14,137\mu\text{m}^3$
GAM	Gamma $n_0(x) = \frac{N_0}{\Gamma(k)\theta^k}x^{k-1}\exp(-x/\theta)$	$N_0 = 100/\text{cm}^3$, $k = 3$ $\theta = 1000\mu\text{m}^3$
BIM	Bimodal Gamma Mixture $n_0(x) = \frac{N_{0,a}}{\Gamma(k_a)\theta_a^{k_a}}x^{k_a-1}\exp(-x/\theta_a)$ $+\frac{N_{0,b}}{\Gamma(k_b)\theta_b^{k_b}}x^{k_b-1}\exp(-x/\theta_b)$	$N_{0,a} = 10/\text{cm}^3$, $N_{0,b} = 100/\text{cm}^3$ $k_a = 4, k_b = 2$ $\theta_a = 1000\mu\text{m}^3, \theta_b = 200\mu\text{m}^3$

Table 3.1: Summary of the three initial conditions tested for collision-only dynamics.

Label	Equation	Duration
1C	Constant kernel collisions $K(x, y) = A$ $A = 10^{-4}/\text{cm}^3/\text{s}$	EXP: 4hr GAM: 4hr BIM: 4hr
1G	Golovin kernel collisions $K(x, y) = B(x + y)$ $B = 1500/\text{s}$	EXP: 2hr GAM: 1hr BIM: 4hr
1H	Hydrodynamic kernel collisions $K(x, y) = C\pi\big(r(x) + r(y)\big)\lvert a(x) - a(y)\rvert$ $C = 10^{-12}\text{cm}^3/\mu\text{m}^4/\text{s}$	EXP: 4hr GAM: 4hr BIM: 4hr
2	Golovin kernel collisions, injection, and removal $K(x, y) = B(x + y)$, $B = 1500/\text{s}$ $I(x) = \frac{x^{k-1}\exp(-x/\theta)}{\Gamma k\theta^k}$, $k = 2, \theta = 200\mu\text{m}^3$ $\dot{P} = 1/\text{s}$, $r_{max} = 25\mu\text{m}$	2hr

Table 3.2: Equations and parameters for the four collision dynamics considered, and the corresponding initial condition/simulation duration pairs tested for each dynamic. For the hydrodynamic case 1H, $r(x)$ and $a(x)$ refer to the radius and area, respectively, of a particle with volume x. Case 2 is initialized as an empty box.

simple closure-based 2- or 3-moment bulk representation cannot capture multiple modes without an additional set of prognostic moments and autoconversion rates; therefore, this test case highlights the information gained from using a more flexible PSD representation.

The dynamics considered (table 2) fall into two categories: (1) collision-coalescence only with three different collision kernels, or (2) Golovin collisions and additional dynamics of particle injection and removal from the box. Cases 1C, 1G, and 1H use a a constant collection efficiency $E_c = 1$ and three collision kernels of increasing complexity: (1C) a constant rate of collision, (1G) a Golovin linear kernel, and (1H) a hydrodynamic kernel where $r(x)$ and $a(x)$ represent the radius and area of a particle of mass x, respectively. The kernel parameters and simulated duration are selected such that the number concentration of the final droplet spectrum has decreased, but does not approach the single-large-particle limit of coalescence-only dynamics. Test case 2 begins with an initially empty box, using the same collision kernel and parameters as 1G, plus a constant prescribed injection rate and injection PSD to mimic entrainment/activation of new particles, and a maximum particle size $r_{max} = 25\ \mu$m (as in Khairoutdinov and Kogan, 2000) for removal/sedimentation of particles approaching a drizzle size-threshold. This set of particle dynamics will drive the PSD to a steady state in which particles enter the system, collide, grow, and precipitate out of the system. While modeling collision-coalescence by itself is a useful numerical test, it requires that the microphysics scheme be able to represent arbitrarily large particles with an accelerating rate of growth. Using a simplified proxy for the introduction of small droplets and removal of large droplets allows for the study of a steady-state PSD.

For the dynamics and initial conditions described above, we investigate the particle size spectra pre- and post-collisions through the marginal mass distribution:

$$\frac{dm}{d\ln(r)} = 3x^2 n(x), \tag{3.7}$$

where dm is the mass of particles in a size range of $d\ln(r)$ in a logarithmic space of spherical particle radius r, corresponding to particle mass x. An L_2 spectral error E_{L2} is calculated as a normalized sum of squared differences for this quantity between the approximated profiles and a reference solution from Lagrangian microphysics:

$$E_{L2} = \frac{\int_0^\infty \left(\frac{dm}{d\ln(r)}_{\text{approx}} - \frac{dm}{d\ln(r)}_{\text{ref}}\right)^2 d\ln(r)}{\int_0^\infty \left(\frac{dm}{d\ln(r)}_{\text{ref}}\right)^2 d\ln(r)}. \tag{3.8}$$

In addition, we consider the first three moments of the PSD, which correspond to total number density, total mass density, and radar reflectivity; these are standard quantities tracked in both climate modeling and weather prediction. Finally, to understand the ability of the BF method to represent the growth of cloud droplets into rain-range droplets, we compute the mass of droplets in the box exceeding a size threshold of $r_{\max} = 25\ \mu$m (Khairoutdinov & Kogan, 2000) as the system evolves. This exceedance can be considered a proxy for precipitation, even though all mass remains in the box and these large particles may continue to collect smaller droplets. (A threshold of 25 μm was shown to have the best performance in distinguishing rain and cloud modes in a different exploratory microphysical scheme (Igel et al., 2022), although the use of a fixed size cutoff is inherently arbitrary.) Although the terminal velocity of a 25 μm particle is insufficient to sediment in a realistic convective updraft, we introduce this threshold as a means of evaluating mass growth in the tail of the particle size distribution, which corresponds to autoconversion in bulk microphysics schemes and relates to the mass of droplets that experience the highest rates of collision-coalescence.

For comparison with the BF method, we solve each test case numerically using the flux method for spectral bin microphysics with 32 single-moment bins (Bott, 1998), a two- or three-moment closure method of moments (Bieli et al., 2022), and a Lagrangian particle-based code called PySDM (v2.5) (Bartman et al., 2022). The bin method used follows the original setup from Bott, 1998, spanning a range of 0.633 μm to 817 μm radius with mass doubling between bins, and a time step selected to be sufficiently small as to prevent numerical instability (1 s to 100 s depending on the dynamics). The choice of 32 bins is common in LES application and contrasts the performance of the BF method with a spectral representation that is too expensive for existing GCMs. The MOM representation uses 2 moments with an exponential closure for the EXP initial condition test cases, and a 3 moment gamma closure for the GAM and BIM initial conditions. The MOM is initialized to match the initial moments of each distribution exactly, and therefore is an exact match of the initial PSD for the EXP and GAM cases. As a high-fidelity reference for the collision-only dynamics, we use the results of Lagrangian microphysics. The collision kernels used in this Lagrangian case have exactly the same functional forms as those used for the collocation, bin, and bulk methods with a constant collection efficiency $E_c = 1$.

These Lagrangian PySDM simulations use $2^{15} = 32,768$ superdroplets to represent the particle population in a box of volume 1 m^3, and are taken as the reference profile for the purposes of discussion and L_2 spectral error. While operational SDM simulations typically use closer to 256 superdroplets per gridbox to study broader microphysical processes, the choice of a sufficiently large quantity of superdroplets was shown to reproduce analytic and numerical solutions for similar collision dynamics in a 0D box setting well (Shima et al., 2009). Furthermore, it is expected that the solution error for the other methods considered in these experiments will outpace that of the Lagrangian results given the large number of superdroplets used to represent the PSD. Thus although the superdroplet method may incur significant errors due to an under-resolution of the size spectrum when fewer superdroplets are used, the PySDM results presented in this work are a reasonable benchmark against which to validate other methods.

The BF method is demonstrated in both a low-complexity and moderate-complexity configuration, with 8 or 16 CSLBF1 basis functions, respectively. Both configurations span a particle size range of 2 μm to 200 μm radius, which corresponds to 26 of the 32 bins used in the flux method. Collocation points are logarithmically spaced over particle volumes corresponding to this size range (particles are assumed spherical with liquid water density). BF shape parameters θ_k are chosen such that the basis functions overlap with their nearest neighbors: $\theta_k = \mu_k - \mu_{k-2}$ and $\theta_1 = \theta_2 = \mu_2$. The method is implemented in the Julia programming language and uses a variable time-step with the DifferentialEquations.jl package (Rackauckas & Nie, 2017). The constrained optimization in equation (3.4) is solved using NonNegLeastSquares.jl v0.4.0 (non-negative least squares). Numerical integrals are computed using Cubature.jl v1.5.1.

3.4 Results

3.4.1 Unimodal collision-coalescence

For the collision-coalescence box cases with a single particle size mode (EXP and GAM), we find that the collocation method with 16 basis functions is able both to reproduce the initial PSD well and to accurately predict the PSD following collisions. The spectra for case 1G-GAM in Figure 3.2 reveal that more than eight basis functions are necessary for this particular BF configuration to approximate the initial condition's primary size mode, while 16 BFs are sufficient to capture the height and location of this mode. The 16-BF configuration outperforms the MOM in predicting the final spectrum, as the primary mode in the MOM is slightly too large and the spectrum too narrow compared with Lagrangian, bin, and BF approach.

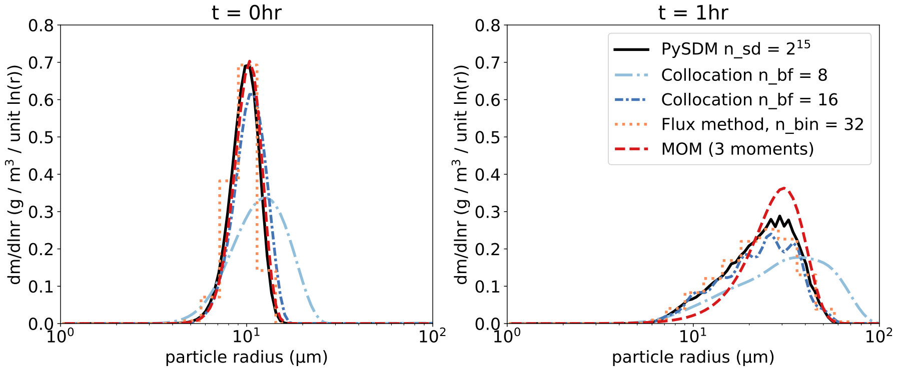


Figure 3.2: Initial spectrum GAM (left) and post-collision spectrum (right) resulting from a Golovin kernel collision-coalescence (1G) for bulk (MOM), bin (flux), and Lagrangian methods, and using the BF collocation method with 8 or 16 degrees of freedom.

The 8-BF configuration artificially broadens both the initial and final PSD due to the wide basis functions required to span the particle size range of interest. This result supports the claim that the BF approach is suitable as a moderate complexity representation of microphysics, whereas a bulk method with closure is appropriate when it is desired to use only a few degrees of freedom.

The results of all three initial conditions and collision kernels are summarized by the L_2 spectral errors of the post-collision spectra in Figure 3.3. The collocation method with 8 BFs performs on par with the MOM in many cases, suffering from broadening of the spectrum particularly in the BIM case, which requires high resolution in the particle size space to capture both narrow modes. With 16 BFs, the BF method outperforms a bulk method in predicting the post-collision spectra in most cases, and it performs similarly to a bin method (with superior accuracy in half of the cases) despite requiring fewer degrees of freedom than the bin approach to span the same particle size range. Errors in the bin representation range from 15–30% relative to the Lagrangian results and may be attributed in part to the stepwise PSD representation. Likewise, the BF approach with 16 degrees of freedom results in errors consistently less than 25% for all three collision kernels investigated with unimodal initial conditions, with some of this error attributable to an under-resolution of the smooth distribution. In contrast, spectral errors in the MOM vary from case-to-case according to the complexity of the collision kernel and initial condition. These results demonstrate the potential for the collocation method to consistently resolve realistic droplet spectra at a level of complexity between that of a bulk method and a 32-single-moment bin scheme.

Next we investigate bulk quantities predicted by each method in figures 3.4 and 3.5, which illustrate the time evolution of the first three moments and exceedance mass, respectively, for the 1G-GAM case, which displays representative behavior among all unimodal test cases. The bulk method of moments outperforms the BF method in predicting the time evolution of the PSD moments, as the first two moments are predicted analytically, and the gamma closure approximation is only employed in computing the second moment. (Although these results indicate that the bulk MOM performs well on all bulk quantities relative even to a bin representation, bulk methods do not typically represent precipitation through exceedance as is done here, but rely on autoconversion parameterizations which introduce significant uncertainty into the model.) The BF method does not exactly conserve mass, especially when fewer BFs are used, in part because the use of compactly supported basis functions prevents the representation of particles larger than the support of the basis functions. The distributional representation of a particle population allows for the formation of unphysically large particles in the collision-coalescence equation; therefore, even though the spectral error is relatively low, the inability of the BF to capture such large particles strongly penalizes the higher-order moments, especially relative to a bulk representation with a globally-supported closure assumption. Furthermore, the matrix inversion in equation (3.4) does not guarantee conservation of mass, particularly where the system of equations might be large and ill-conditioned. Despite this challenge and lack of a guarantee, the BF method does a reasonable job recovering mass conservation throughout the simulation time, with mass variations (relative to the exact initial mass and reference Lagrangian solution, which is mass-conserving) of up to 17% in the 8-BF case due to an early overprediction of mass, and up to a loss of 9% of total mass in the 16-BF case, with the maximum error incurred near the end of the simulation. Among all unimodal test cases summarized in Figure 3.3, mass error for the 16-BF case is most severe for case 1H-GAM (not shown) with an initial overprediction of 17%.

The second moment is overestimated by the 8-BF method initially due to error in projecting the initial PSD onto the basis space: the initial projection slightly overpredicts the size of some droplets, but not so much as to miscategorize them in the exceedance regime in either BF case, as indicated in figures 3.2 and 3.5. The results further indicate that the 8-BF representation is not sufficient to accurately represent the short-time dynamics of the moment evolution, as evidenced by jumps

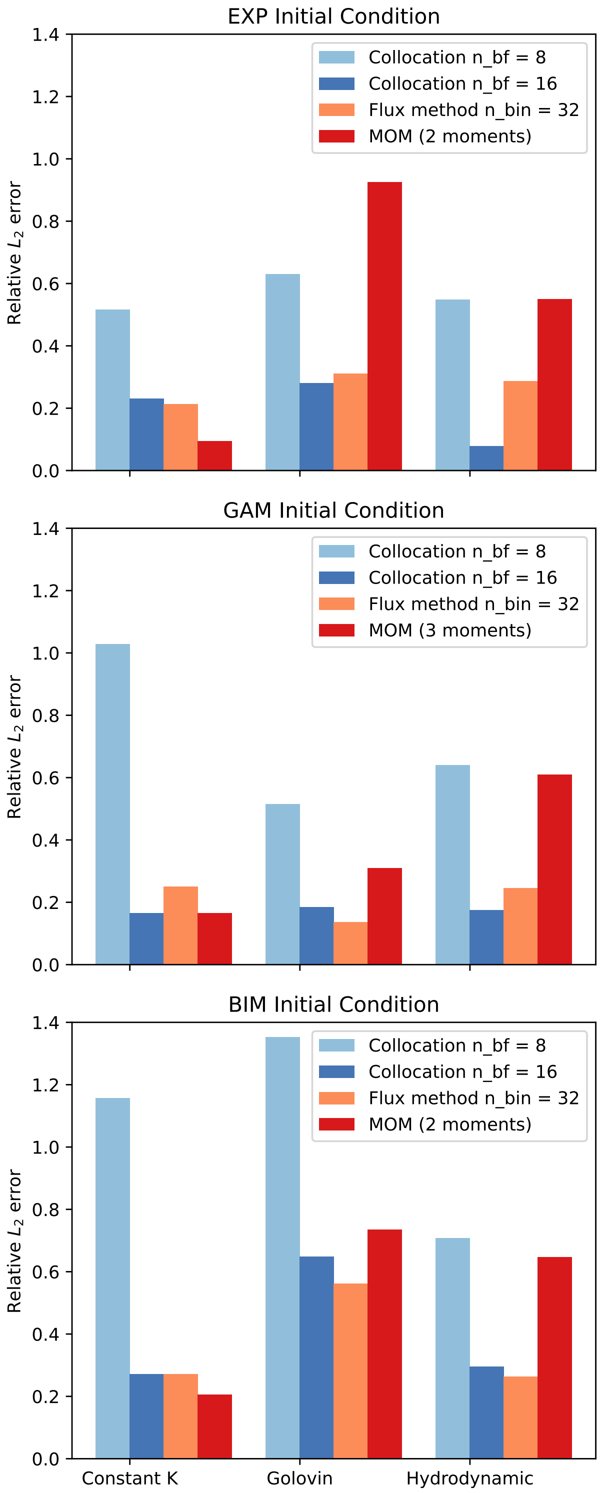


Figure 3.3: Spectral error (L_2) for the bulk, bin (flux), and BF methods with 8 or 16 basis functions, computed relative to a Lagrangian PySDM result. Errors are computed for each of three coalescence-only experiments (case 1C, 1G, and 1H, respectively; bar colors), and each of three initial conditions (EXP, GAM, BIM; top to bottom).

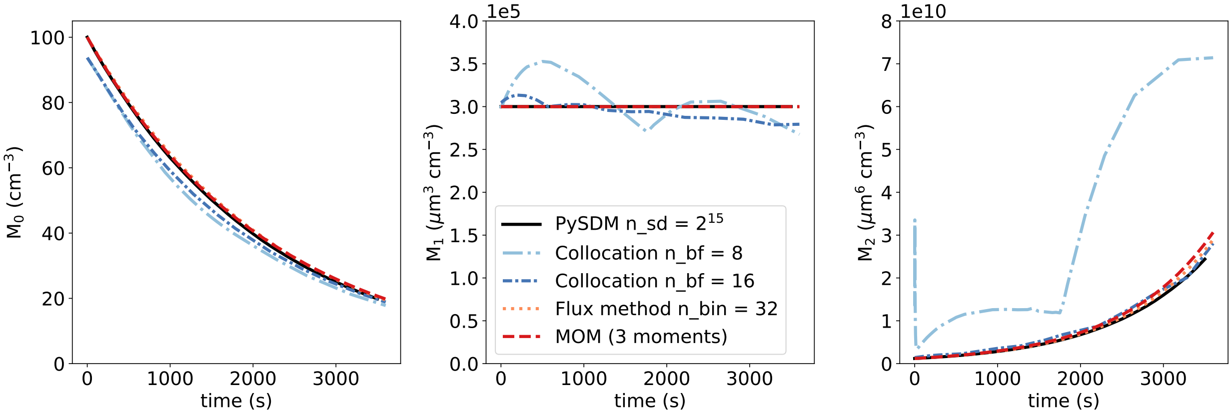


Figure 3.4: Evolution of the first three moments (left to right) of the GAM initial condition over time for bulk, bin, and BF method with 8 or 16 degrees of freedom using Golovin collision dynamics (1G).

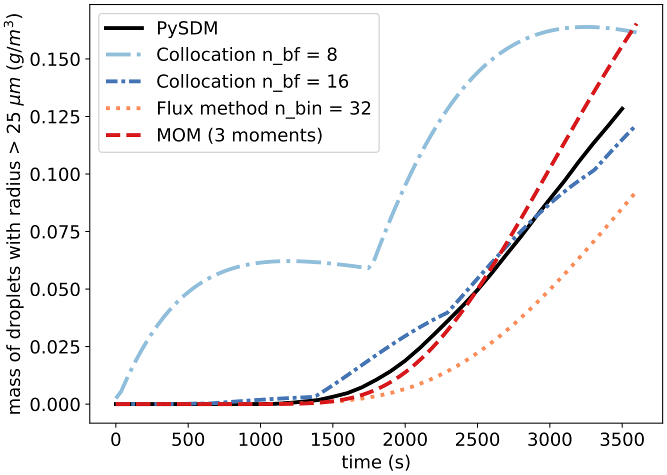


Figure 3.5: Mass density of droplets exceeding 25μm in radius for Lagrangian, bulk, bin, and collocation methods as a function of time for Golovin collision dynamics (1G) and initial condition GAM.

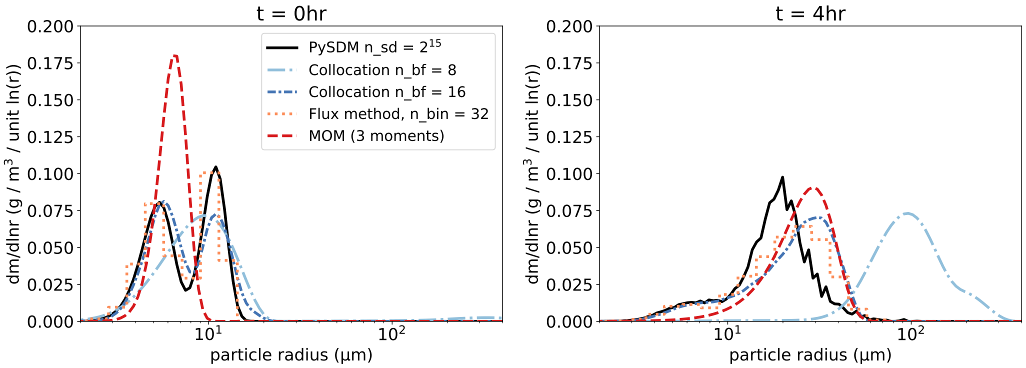


Figure 3.6: Spectra following collision-coalescence of a bimodal droplet population BIM using Lagrangian, bulk, bin, and BF methods with a Golovin kernel (case 1G).

in the first and second moments as well as the exceedance mass. As noted in the spectral case, the necessity for the basis functions to span the entire size range leads to an artificial broadening of the spectrum in this case, as well as to unphysical jumps in bulk quantities in time as soon as a basis function corresponding to a larger size mode is first activated. This challenge is alleviated by using more BFs (still lower complexity than the bin representation), and, in fact, the 16-BF representation outperforms the bin method in predicting exceedance mass (Figure 3.5): the tail of the bin representation is underresolved with only 32 bins, leading to an underprediction of the mass of large precipitation-range droplets. Indeed, despite shortcomings in predicting PSD moments, the BF method improves prediction of the second moment over the bin method in 6 out of 9 test cases (not shown), indicating that the method still performs well at capturing the mass of particles which lie in the tail of the distribution. This indicates a powerful capability for this flexible spectral method to accurately capture the rate of transition from small cloud droplets to a coalescence-driven "rain" mode without prior assumptions of a size threshold or autoconversion rate, while also continuing to resolve the size spectrum of the rain-mode droplets.

3.4.2 Multimodal collision-coalescence

One strength of the BF method is its ability to represent up to n_{BF} modes of a PSD, where n_{BF} is the number of basis functions used. By contrast, bulk methods can represent at most one droplet mode, and bin methods lose spectral detail of the modes due to the piecewise constant representation of the PSD. We demonstrate in Figure 3.6 the test case 1G-BIM: Golovin collision-coalescence with an initially bimodal

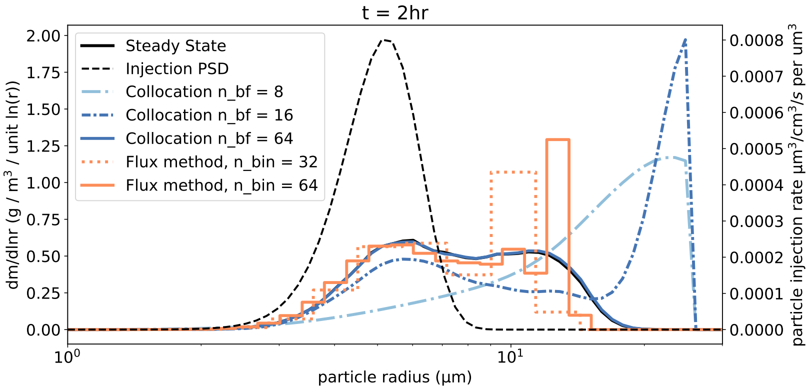


Figure 3.7: Steady state PSD for the third case with collisions, sedimentation, and injection, using a bin method with 32 or 64 bins, the BF method with 8, 16, or 64 basis functions, and a steady state solution computed using basis functions and Newton's method. The PSD of injected particles is plotted as a dashed black line with units on the right y axis.

distribution. The Lagrangian, bin, and 16-BF results indicate that the smaller PSD mode mostly disappears after 4 hours, leaving behind only a small shoulder in the PSD and broadening and shifting the mode toward larger particles. With 16 degrees of freedom, the BF method accurately captures both of these modes during the PSD evolution, while a large initial condition error in the 8-BF case propagates to an artificially large final PSD. By comparison, the gamma-closure MOM cannot represent the initial or final PSD exactly due to the underlying unimodal closure assumption, but it only slightly overpredicts the size of the dominant mode in the post-collision spectra. The bin method accurately predicts droplets in both size ranges, although the underlying piecewise-constant representation under-resolves the complexity of the multimodal distribution initially.

3.4.3 Collision-coalescence with injection and removal

When including removal of large particles and introduction of small particles (case 2), we investigate the steady-state PSD and the time evolution of the PSD moments to a steady state. No Lagrangian or MOM results are presented in this case, as the removal and injection process rates used are not applicable in those frameworks. As a high resolution reference, we instead present results from a 64-bin representation and a 64-BF representation over the same size range, and a steady-state numerical solution. This numerical steady state is computed with a 64-BF representation by setting $d_t\tilde{\mathbf{m}}(t) = 0$ in equation (3.3) and iterating using Newton's method with an

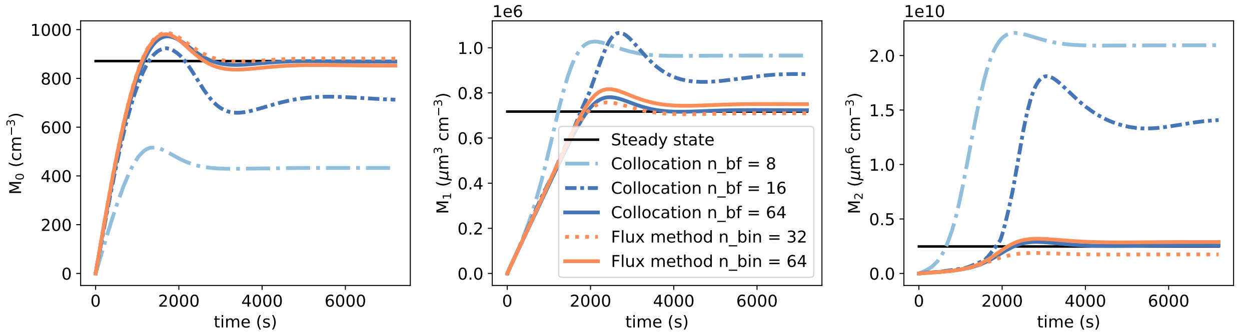


Figure 3.8: Time-series evolution of first three moments of the distribution for the collocation and bin methods with collisions, precipitation, and injection (case 2), as well as high-fidelity and steady state reference values.

initial guess of $\mathbf{P_{inj}}$, the projection of the injection PSD $I(x)$ (table 1) on the basis space.

In Figure 3.7, we see that the steady state and 64-BF results indicate that the injected PSD should broaden and form a secondary mode at around 10 μm in size. Particles enter the system, grow through collisions, and exit once they reach 25 μm in size. However, this depletion of large particles is reflected by a substantial decrease in slightly smaller particles, down to 10 μm radius. This can be explained by the fact that particles larger than 10 μm are (a) far more likely to collide with each other, given the Golovin kernel; and (b) likely to form a particle larger than 25 μm upon coalescence and thus leave the system. The bin method reflects this trend while overpredicting the mass of droplets in this 10–12 μm size range, yet it still captures the first three moments of the distribution (Figure 3.8) relatively well, with only marginal improvement from doubling the number of bins. In contrast, the BF method does a poor job in this instance, particularly when only 8-BF are used: because the 8-BF approach artifically broadens the injected spectrum, as seen in previous spectral results, it similarly artificially accelerates the collisions toward larger droplets. The truncation of the integrals at the exceedence size of 25 μm is not sufficient to capture the depletion of slightly smaller particles in either BF case, as the distribution peaks near 25 μm in both BF cases. This large-particle peak is reflected in the overprediction of the steady-state mass (with errors in total mass up to 30%) and second moment in Figure 3.8, though the transient growth period up to around 2000 s is well represented when 16 BFs are used.

This set of particle dynamics, particularly removal of large particles using a size threshold, demonstrates that the BF method is ill-suited to represent dynamics with fixed size-cutoffs at long time scales, relative to a traditional bin scheme. However, the coalescence-only experiments presented earlier indicate that the BF method performs well in predicting the tail of the PSD when these large droplets are not removed from the system, indicating that it is capable of representing cloud and rain droplets as a continuous spectrum, rather than using separate prognostic variables as is typically done in bulk methods.

3.4.4 Computational Complexity

The BF method offers similar computational scaling to a bin spectral method, but is higher complexity than a traditional multi-moment bulk method. Bulk methods with a closure assumption scale with the number of moments, $O(N_{\text{mom}})$ when the relationship between the prognostic moments and PSD parameters is known, but more complex PSD closures may require nonlinear operations, tabulations, or nonlinear optimization, leading to computationally intensive operations at each time step. Spectral bin methods such as the flux method used here (Bott, 1998) scale quadratically with the number of bins, $O(N_{\text{bin}}^2)$, as each pair of bins is considered sequentially for the coalescence dynamic. The basis function method scales either cubically or quadratically depending on the choice of basis (see appendix B). While the initial precomputation for the BF method is cubic in the number of basis functions, a compactly supported basis will lead to quadratic operations in the forward time-marching of equation (3.6), as the third-order tensor $\mathbf{Q}$ is sparse. This places the BF method at the same order of complexity as other spectral methods, $O(N_{\text{BF}}^2) \sim O(N_{\text{bin}}^2)$, but with performance that can meet or exceed bin scheme accuracy with half or fewer of the required degrees of freedom. Thus the collocation method could recover bin-like accuracy with only 1/4 of the computational cost.

3.5 Discussion and Conclusions

This paper describes and demonstrates a novel method to represent the particle size distribution of droplets for warm-rain atmospheric microphysics. Collocation of basis functions provides a more flexible PSD approximation than either bin microphysics or the method of moments with closure (bulk microphysics). In particular, selecting BFs which are themselves distributions generalizes traditional spectral bin methods to a smoothed representation that can be interpreted as the sum of droplet size modes, and the collocation approach applies to any choice of basis function form, unlike the bin approach, where the numerical methods are specific to a piecewise constant representation. The method is also appropriate for applications where more than three degrees of freedom (the most usually provided in a bulk scheme) are desired, but where full bin complexity is infeasible. In this low-complexity limit, collocation of basis functions can be considered a linear closure relating the mass density at the collocation points to a BF weight vector.

Tested in a variety of box model settings, we find that the BF method improves spectral accuracy under collision-coalescence dynamics compared to a three-moment bulk method, while using fewer degrees of freedom than a bin method. The spectral detail from the BF approach allows for a precise calculation of mass in the tail of the distribution (exceedance), which could avert the need for precipitation parameterizations that are required by bulk methods. Another strength of the method is its ability to represent multimodal distributions, unlike 3-moment bulk methods. Under dynamics including injection and removal which drive the PSD toward a steady state, the BF method does well in predicting the size distribution and moments in the transient regime, but overpredicts the quantity of large particles in the steady state size distribution. This finding indicates that the BF method is better suited to continuous dynamics (rather than size thresholding), and therefore might perform well in a context with no artificial separations between cloud and rain droplets.

In general, the BF method is a more flexible framework than bulk or bin methods: the suggested implementation can receive an arbitrary set of microphysical processes and automatedly perform all required numerical integrations. This is in contrast to bin methods, which require tabulated collision and breakup kernels that are dependent on the bin discretization, and in contrast to bulk methods, which frequently include hard-coded parameterizations and closures. The ability to specify arbitrary functional process rates for the BF method will be especially useful for reducing microphysics parameter uncertainty in atmosphere models. While this study focuses on the collisional-coalescence dynamic, other warm rain dynamics such as condensation-evaporation could be formulated as processes driving the collocation weights in time, as in equation (3.6), though testing additional warm rain dynamics is beyond the scope of this work. Condensation-evaporation is a source of artificial spectral broadening in typical bin methods (Khain et al., 2015) unless movable bins are implemented. Due to the high cost of recomputing numerical integrals for coalescence and other dynamics when collocation points are updated, we would recommend fixing collocation points in particle size space throughout time. The simplest approach would simply substitute the collocation representation and points into the diffusional growth equation, performing automatic differentiation of the basis functions, and adding a new variable for supersaturation to the coupled set of ODEs as in equation (3.6) plus resulting source terms to the right hand side of the equation. As this approach requires no remapping over the particle mass space, it might avert some issues related to numerical diffusion experienced in bin schemes (while potentially incurring other penalties, such as mass conservation).

The BF method does have limitations. First, although the linear system in equation (3.4) is solved in mass density space with a positivity constraint, the method does not exactly conserve mass for collision-coalescence-only dynamics. When employed with compactly supported basis functions, the method can only represent particles up to a maximum size, unlike bulk or Lagrangian methods. This shortcoming manifests in errors in the higher order moments of the PSD, including some mass loss from the system (Figure 3.4). Some potential solutions could involve allowing for globally supported basis functions at the tail of the distribution, or periodically rescaling the weight vector to exactly conserve mass in the system. Alternatively, a mass conserving constraint could be imposed at each time step in addition to the positivity constraint, although doing so could result in numerical instability. As this mass non-conservation typically only appears at the long-time limit of the simulation when the total number concentration of particles is substantially depleted, we anticipate that it would not contribute a dominant source of error on short (sub-hourly) time scales when a full set of microphysical dynamics and flow field computations are included. Instead, a simple rescaling of the basis function weights before advecting microphysical quantities across grid boxes or computing saturation adjustments would be sufficient to close the mass balance without strongly impacting the spectral results. (For instance, this technique would require less than a 10% rescaling in the 16-BF case presented in Figure 3.4.) As indicated by the coalescence-injection-removal case in Figure 3.8, rescaling is likely unnecessary for mass conservation on short time scales, but would be straightforward on longer time scales where a steady state in the PSD is approached. Alternative approaches from the aerosol modeling community such as the Quadrature-based Method of Moments McGraw, 1997 do not suffer these mass conservation challenges, and therefore could be an advantageous alternative to collocation.

When particle removal and injection are considered, the method is able to accurately predict moments in the transient regime, and in fact over-predicts the quantity of large particless. Therefore, additional testing with a complete set of microphysical processes and a comprehensive range of initial conditions will be required to determine whether further refinement is necessary. Future work to improve and test this novel microphysics method will involve incorporating additional microphysical processes such as diffusional growth, as well as employing one, two, and three-dimensional simulations to test the ability of the method to reproduce mesoscale cloud properties and to assess the severity of mass nonconservation without imposing additional constraints. Further testing of the method in a one-dimensional setting with spatial advection as well as diffusional growth will also be necessary to assess how susceptible the collocation implementation is to numerical diffusion, as is often observed with bin schemes.

The BF method presented here and tested for coalescence dynamics improves spectral accuracy over bulk methods with fewer degrees of freedom than a bin method, and it has the potential to reduce the computational cost of microphysics even further. Using inspiration from proposed moving bin schemes, the locations or shapes of BFs could be automatically selected and periodically updated to maximize the information potential provided by only n_{BF} degrees of freedom. While this approach could impose the cost of recomputing numerical integrals, if done sparingly and intelligently it would cluster basis functions near the most-weighted droplet modes, improving the accuracy-complexity tradeoff. Another potential benefit of the collocation representation is the ability to use multidimensional basis functions: one independent variable could be the droplet size, as in this work, while other particle properties such as aerosol hygroscopicity, ice riming fraction, or surface tension could occupy additional inputs. This multidimensional representation has been explored for aerosol bin schemes (Lebo & Seinfeld, 2011), as well as for ice bulk methods (Morrison & Milbrandt, 2015). However, it may be more computationally efficient to represent multiple particle properties in the BF framework due to the flexibility of selecting radial basis functions with compact support to generate a sparse system and lessen the computational burden. Such a representation could eliminate the uncertainties of conversion parameterizations and of information loss from aggregating particles into categories with distinct sets of microphysical dynamics. This potential lights a path toward unifying the numerical representation of all microphysical particles in a single, consistent framework.

3.6 Appendix A: Basis functions, collocation points, and hyperparameters

The BF collocation parameters demonstrated in this study are briefly explained. As the collocation points correspond to the droplet mode represented by each BF, we should not assume a priori any particular initial or final distribution of particles. However, we can use the inherent length scales of the physical system to aid the setup. For cloud droplets and aerosols, the size domain should extend from $x_{\min} \geq 0\mu$m to the size of the largest particles $x_{\max}$ that do not sediment out of the system or instantaneously break up, hence making a finite domain approximation reasonable. Furthermore, we draw inspiration from bin microphysics to suggest logarithmically spaced collocation points over the domain.

The basis function family and their hyperparameters should then be selected to ensure a few criteria:

1. The entire domain $[0, x_{\max}]$ is spanned with some minimum probability.
2. There should be no particles with negative or infinite mass; that is, $\phi_k(x < 0), \phi_k(x \to \infty) = 0$ for all basis functions.
3. BF hyperparameters should be selected to minimize oscillations and jumps in the approximated distribution.

The first condition is equivalent to requiring either globally-supported BFs, such that $\phi(x) > 0 \ \forall x$, or sufficient overlap of compactly-supported BFs, which are positive over some interval and zero elsewhere. The second condition cannot be met exactly for any BFs that are globally supported over $(-\infty, \infty)$, therefore we suggest using either compactly-supported BFs (CSBFs) or exponentially decaying BFs. CSBFs are additionally recommended due to their favorable numerical properties: Zhang et al., 2000 demonstrate that CSBFs result in a better conditioned system of equations (as in equation (3.5)). The third criterion is the trickiest and will depend on the family of BFs chosen. As a simple heuristic for a two-parameter family such as Gaussians, we suggest setting the scale factors as some multiple of the spacing between collocation points to ensure support and smoothness over the domain. More sophisticated methods of setting the hyperparameters, such as optimization over a set of potential distributions or constraints on fluctuations in the second derivatives, are possible but beyond the scope of this paper.

Several families of basis functions are suitable to approximate a droplet size distribution, such as Gaussian, gamma, and lognormal distributions. In order to obtain a compactly supported basis, however, we propose to use a version of the CSRBF1, a compactly supported Gaussian approximation proposed by Wu, 1995, modified to instead uses a logarithmic argument. This basis function, which we will refer to as CSLBF1 (compactly supported lognormal BF 1) takes the form:

$$\phi(r) = \begin{cases} \frac{12}{35}(1-r)^4(4+16r+12r^2+3r^3) & r \leq 1 \\ 0 & r > 1 \end{cases} \tag{3.9}$$

with argument

$$r = \frac{|\log(x) - \mu|}{\theta},$$

where μ is the collocation point and θ is a scale factor. Given that CSRBF1 approximates a normal distribution, CSLBF1 approximates a lognormal distribution, which is better suited to particle distributions as it is right skewed.

3.7 Appendix B: Collocation of BFs for the SCE

Evaluating equation (3.1) with arbitrary additional processes $\mathbf{P_l}$ in mass density at collocation point μ_j, we find

$$\begin{aligned} \partial_t \mu_j n(\mu_j, t) =& 1/2\mu_j \int_0^{\mu_j} n(\mu_j - y, t) n(y, t) K(\mu_j - y, y) E(\mu_j - y, y) dy \\ & - \mu_j n(\mu_j, t) \int_0^{x_{\max}-\mu_j} n(y, t) E(\mu_j, y) K(\mu_j, y) dy + \sum_{l=1}^{N_{proc}} P_l(\mu_j, n(\mu_j, t)). \end{aligned} \tag{3.10}$$

Substituting the collocation approximate solution for local mass density, $x\tilde{n}(x,t) = \sum_{k=1}^{p} x\phi_k(x)c_k(t)$, this time derivative becomes

$$\begin{aligned} \partial_t \tilde{m}_j(t) =& 1/2 \sum_{k=1}^{n_{\mathrm{BF}}} \sum_{l=1}^{n_{\mathrm{BF}}} \mu_j c_k(t) c_l(t) \int_0^{\mu_j} \phi_k(\mu_j - y) \phi_l(y, t) K(\mu_j - y, y) E(\mu_j - y, y) dy \\ & - \sum_{k=1}^{n_{\mathrm{BF}}} \sum_{l=1}^{n_{\mathrm{BF}}} \mu_j c_k(t) c_l(t) \phi_k(\mu_j) \int_0^{x_{\max}-\mu_j} \phi_l(y) K(\mu_j, y) E(\mu_j, y) dy + \sum_{l=1}^{N_{proc}} \mu_j P_l(\mu_j, \tilde{n}(\mu_j, t)). \end{aligned} \tag{3.11}$$

The collision-coalescence dynamics are summarized via a third-order tensor in mass density: **Q**, with

$$\begin{aligned} Q_{jkl} =& 1/2\mu_j \int_0^{\mu_j} \phi_k(\mu_j - y) \phi_l(y, t) K(\mu_j - y, y) E(\mu_j - y, y) dy \\ & - \mu_j \phi_k(x\mu_j) \int_0^{x_{\max}-\mu_j} \phi_l(y) K(\mu_j, y) E(\mu_j, y) dy \end{aligned} \tag{3.12}$$

The overall dynamics are then summarized by cubic collision-coalescence dynamics plus the additional processes projected onto the basis space as in equation (3.5) to obtain the terms $\mathbf{P}_l = \big(\mu_1 P_l(\mu_1), \mu_2 P_l(\mu_2), \ldots, \mu_k P_l(\mu_k)\big)$ in equation (3.6).

Many of the quantities in equation (3.6) can be precomputed and stored for a given set of basis functions. These precomputations include:

- The linear system, $\mathbf{\Phi}$;
- The third order tensor $\mathbf{Q}$ which can be computed numerically via quadrature or Monte Carlo integration, given a functional form of the kernel;
- Appropriate projection of additional processes onto the basis space to obtain $\mathbf{P}_l$. For the purpose of ensuring mass conservation, this may require computing the first moments of the basis functions over the integration window $[0, x_{\max}]$;
- The initial condition at the collocation points $\tilde{\mathbf{m}}(0)$.

The computation of $\mathbf{Q}$ scales cubically with the number of collocation points for globally supported basis functions, and quadratically for partially overlapping compactly supported basis functions. The dynamical system in equation (3.6) involves at most cubic vector-tensor multiplication and function evaluations for the tensor-vector inner products, and therefore a small system of basis functions is more likely to be limited by the time-stepping scheme or matrix inversion than by the precomputation. Another advantage of choosing compactly supported basis functions is that the constant-collocation matrix $\mathbf{\Phi}$ can be N-diagonal (CSBF's that only overlap their nearest neighbors will result in a tridiagonal system, for example) thus making the inversion much more computationally efficient. Finally, using CSBFs limits the range of particle sizes to a finite domain, making numerical integration more straightforward.

Data Availability

The implementation of basis function collocation and examples used in this work can be found in the package RBFCloud.jl at https://doi.org/10.5281/zenodo.6984349, or on github at https://github.com/edejong-caltech/RBFCloud.jl. The 3-moment bulk scheme uses the package Cloudy.jl, available at https://github.com/CliMA/Cloudy.jl, and the Lagrangian microphysics package PySDM is available at https://github.com/atmos-cloud-sim-uj/PySDM.

Acknowledgements

We thank Anna Jaruga, Melanie Bieli, Clare Singer, Zach Lebo, and John Seinfeld for feedback, insights, and discussion. Additional thanks go to Jakob Shpund for providing access to the Bott flux method bin implementation. E. de Jong was supported by a Department of Energy Computational Sciences Graduate Fellowship. This research was additionally supported by Eric and Wendy Schmidt (by recommendation of Schmidt Futures) and the Heising-Simons Foundation. Part of the research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004).

References

Andrejczuk, M., Grabowski, W. W., Reisner, J., & Gadian, A. (2010). Cloud-aerosol interactions for boundary layer stratocumulus in the Lagrangian Cloud Model. *Journal of Geophysical Research: Atmospheres*, *115*(22). https://doi.org/10.1029/2010JD014248

Andrejczuk, M., Reisner, J. M., Henson, B., Dubey, M. K., & Jeffery, C. A. (2008). The potential impacts of pollution on a nondrizzling stratus deck: Does aerosol number matter more than type? *Journal of Geophysical Research: Atmospheres*, *113*(D19). https://doi.org/10.1029/2007JD009445

Arakawa, A. (2004). The Cumulus Parameterization Problem: Past, Present, and Future. *Journal of Climate*, *17*(13), 2493–2525.

Bartman, P., Bulenok, O., Górski, K., Jaruga, A., Łazarski, G., Olesik, M. A., Piasecki, B., Singer, C. E., Talar, A., & Arabas, S. (2022). PySDM v1: Particle-based cloud modeling package for warm-rain microphysics and aqueous chemistry. *Journal of Open Source Software*, *7*(72), 3219. https://doi.org/10.21105/joss.03219

Berry, E. X. (1967). Cloud Droplet Growth by Collection. *Journal of Atmospheric Sciences*, *24*(6), 688–701. https://doi.org/10.1175/1520-0469(1967)024<0688:CDGBC>2.0.CO;2

Berry, E. X., & Reinhardt, R. L. (1974). An Analysis of Cloud Drop Growth by Collection: Part I. Double Distributions. *Journal of the Atmospheric Sciences*, *31*(7), 1814–1824.

Bieli, M., Dunbar, O. R. A., de Jong, E. K., Jaruga, A., Schneider, T., & Bischoff, T. (2022). An Efficient Bayesian Approach to Learning Droplet Collision Kernels: Proof of Concept Using "Cloudy," a New n-Moment Bulk Microphysics Scheme. *Journal of Advances in Modeling Earth Systems*, *14*(8), e2022MS002994. https://doi.org/10.1029/2022MS002994

Bott, A. (1998). A Flux Method for the Numerical Solution of the Stochastic Collection Equation. *Journal of the Atmospheric Sciences*, *55*(13), 2284–2293.

Fan, J., Wang, Y., Rosenfeld, D., & Liu, X. (2016). Review of Aerosol–Cloud Interactions: Mechanisms, Significance, and Challenges. *Journal of the Atmospheric Sciences*, *73*(11), 4221–4252.

Filbet, F., & Laurençot, P. (2004). Numerical Simulation of the Smoluchowski Coagulation Equation. *SIAM Journal on Scientific Computing*, *25*(6), 2004–2028.

Franke, C., & Schaback, R. (1998). Solving partial differential equations by collocation using radial basis functions. *Applied Mathematics and Computation*, *93*(1), 73–82.

Gelbard, F., & Seinfeld, J. H. (1978). Numerical solution of the dynamic equation for particulate systems. *Journal of Computational Physics*, *28*(3), 357–375.

Gettelman, A., Gagne, D. J., Chen, C.-C., Christensen, M. W., Lebo, Z. J., Morrison, H., & Gantos, G. (2021). Machine Learning the Warm Rain Process. *Journal of Advances in Modeling Earth Systems*, *13*(2), e2020MS002268.

Ghan, S. J., Abdul-Razzak, H., Nenes, A., Ming, Y., Liu, X., Ovchinnikov, M., Shipway, B., Meskhidze, N., Xu, J., & Shi, X. (2011). Droplet nucleation: Physically-based parameterizations and comparative evaluation. *Journal of Advances in Modeling Earth Systems*, *3*(4).

Igel, A. L., Morrison, H., Santos, S. P., & van Lier-Walqui, M. (2022). Limitations of Separate Cloud and Rain Categories in Parameterizing Collision-Coalescence for Bulk Microphysics Schemes. *Journal of Advances in Modeling Earth Systems*, *14*(6), e2022MS003039. https://doi.org/10.1029/2022MS003039

Igel, A. L. (2019). Using an Arbitrary Moment Predictor to Investigate the Optimal Choice of Prognostic Moments in Bulk Cloud Microphysics Schemes. *Journal of Advances in Modeling Earth Systems*, *11*(12).

Intergovernmental Panel on Climate Change. (2014). *Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [doi: 10.1017/CBO9781107415324]. Cambridge University Press.

Kessler, E. (1969). On the Distribution and Continuity of Water Substance in Atmospheric Circulations [doi: 10.1007/978-1-935704-36-2_1]. In E. Kessler (Ed.), *On the Distribution and Continuity of Water Substance in Atmospheric Circulations* (pp. 1–84). American Meteorological Society.

Khain, A. P., Beheng, K. D., Heymsfield, A., Korolev, A., Krichak, S. O., Levin, Z., Pinsky, M., Phillips, V., Prabhakaran, T., Teller, A., Heever, S. C. v. d., & Yano, J.-I. (2015). Representation of microphysical processes in cloud-resolving models: Spectral (bin) microphysics versus bulk parameterization. *Reviews of Geophysics*, *53*(2), 247–322.

Khairoutdinov, M., & Kogan, Y. (2000). A New Cloud Physics Parameterization in a Large-Eddy Simulation Model of Marine Stratocumulus. *Monthly Weather Review*, *128*(1). https://doi.org/10.1175/1520-0493(2000)128<0229:ANCPPI>2.0.CO;2

Lebo, Z. J., & Seinfeld, J. H. (2011). A continuous spectral aerosol-droplet microphysics model. *Atmospheric Chemistry and Physics*, *11*(23), 12297–12316.

McGraw, R. (1997). Description of Aerosol Dynamics by the Quadrature Method of Moments, *Aerosol Science and Technology*, *27*. https://doi.org/10.1080/02786829708965471

Milbrandt, J. A., & Yau, M. K. (2005). A Multimoment Bulk Microphysics Parameterization. Part I: Analysis of the Role of the Spectral Shape Parameter. *Journal of the Atmospheric Sciences*, *62*(9), 3051–3064.

Morrison, H., & Grabowski, W. W. (2007). Comparison of Bulk and Bin Warm-Rain Microphysics Models Using a Kinematic Framework. *Journal of the Atmospheric Sciences*, *64*(8), 2839–2861.

Morrison, H., & Grabowski, W. W. (2008). Modeling Supersaturation and Subgrid-Scale Mixing with Two-Moment Bulk Warm Microphysics. *Journal of Atmospheric Sciences*, *65*(3), 792–812.

Morrison, H., & Milbrandt, J. A. (2015). Parameterization of Cloud Microphysics Based on the Prediction of Bulk Ice Particle Properties. Part I: Scheme Description and Idealized Tests. *Journal of the Atmospheric Sciences*, *72*(1), 287–311.

Morrison, H., van Lier-Walqui, M., Kumjian, M. R., & Prat, O. P. (2019). A Bayesian Approach for Statistical–Physical Bulk Parameterization of Rain Microphysics. Part I: Scheme Description. *Journal of the Atmospheric Sciences*, *77*(3), 1019–1041.

Morrison, H., van Lier-Walqui, M. v., Fridlind, A. M., Grabowski, W. W., Harrington, J. Y., Hoose, C., Korolev, A., Kumjian, M. R., Milbrandt, J. A., Pawlowska, H., Posselt, D. J., Prat, O. P., Reimel, K. J., Shima, S.-I., Diedenhoven, B. v., & Xue, L. (2020). Confronting the Challenge of Modeling Cloud and Precipitation Microphysics. *Journal of Advances in Modeling Earth Systems*, *12*(8). https://doi.org/10.1029/2019MS001689
e2019MS001689 2019MS001689.

Rackauckas, C., & Nie, Q. (2017). DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia. *Journal of Open Research Software*, *5*(1), 15.

Randall, D., Khairoutdinov, M., Arakawa, A., & Grabowski, W. (2003). Breaking the Cloud Parameterization Deadlock. *Bulletin of the American Meteorological Society*, *84*(11), 1547–1564.

Riechelmann, T., Noh, Y., & Raasch, S. (2012). A new method for large-eddy simulations of clouds with Lagrangian droplets including the effects of turbulent collision. *New Journal of Physics*, *14*(6), 065008. https://doi.org/10.1088/1367-2630/14/6/065008

Rodríguez Genó, C. F., & Alfonso, L. (2022). Parameterization of the collision–coalescence process using series of basis functions: COLNETv1.0.0 model development using a machine learning approach. *Geoscientific Model Development*, *15*(2), 493–507.

Seifert, A., & Beheng, K. D. (2006). A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 1: Model description. *Meteorology and Atmospheric Physics*, *92*(1-2), 45–66.

Shima, S.-I., Kusano, K., Kawano, A., Sugiyama, T., & Kawahara, S. (2009). The super-droplet method for the numerical simulation of clouds and precipitation: A particle-based and probabilistic microphysics model coupled with a non-hydrostatic model. *Quarterly Journal of the Royal Meteorological Society*, *135*(642), 1307–1320. https://doi.org/10.1002/qj.441

Stephens, G. L., L'Ecuyer, T., Forbes, R., Gettelmen, A., Golaz, J.-C., Bodas-Salcedo, A., Suzuki, K., Gabriel, P., & Haynes, J. (2010). Dreary state of precipitation in global models. *Journal of Geophysical Research: Atmospheres*, *115*(24).

Tzivion (Tzitzvashvili), S., Feingold, G., & Levin, Z. (1987). An Efficient Numerical Solution to the Stochastic Collection Equation. *Journal of Atmospheric Sciences*, *44*(21), 3139–3149.

Wu, Z. (1995). Compactly supported positive definite radial functions. *Advances in Computational Mathematics*, *4*(1).

Young, K. C. (1974). A Numerical Simulation of Wintertime, Orographic Precipitation: Part I. Description of Model Microphysics and Numerical Techniques. *Journal of the Atmospheric Sciences*, *31*(7), 1735–1748.

Zhang, X., Song, K. Z., Lu, M. W., & Liu, X. (2000). Meshless methods based on collocation with radial basis functions. *Computational Mechanics*, *26*(4), 333–343.

Chapter 4

AVOIDING CONVERSION RATES WITH FLEXIBLE MOMENT-BASED MICROPHYSICS IN CLOUDY.JL

Abstract

Modern climate models represent warm-phase hydrometeors by separating liquid droplets into cloud and rain categories and tracking one, two, or three statistical moments of their distributions. Such schemes require parameterizations to represent rates of conversion of cloud droplets into rain droplets, replacing a physically interpretable process (growth by coalescence based on droplet sizes) with artificial rates of autoconversion and accretion. This work presents an alternative moment-based approach in which an arbitrary number of subdistributions may be used to represent the overall droplet size spectrum, including cloud and rain. Transfer of moments between these subdistributions is computed from explicit integration over pairs of subdistributions, akin to the superdroplet method. This method eliminates the need for conversion rates while also avoiding issues of non-uniqueness when converting between moments and parameters of a multimodal distribution. We demonstrate the accuracy of this approach in modeling rain initiation and evolution of the droplet spectra in comparison to standard bulk parameterizations and a high-fidelity superdroplet method. This self-consistent and flexible microphysics framework offers an efficient and accurate alternative to conventional microphysics schemes.

Plain Language Summary

Most climate models rely on inflexible approximate equations to model cloud and rain droplets in the atmosphere. These equations make unnecessary and sometimes false assumptions about the sizes and physics of these droplets, leading to model errors. Our work presents an alternative mathematical framework that eliminates the need to distinguish between "cloud", "rain", or other cloud particle categories. This method shows comparable accuracy to much more complex and expensive cloud droplet simulations, at significantly lower computational cost.

4.1 Introduction

Beginning in the 1960s, climate modelers developed simple yet enduring representations of cloud microphysics that track only one or a few bulk quantities of a large and diverse hydrometeor population (Kessler, 1969). These single-moment and two-moment schemes distill the complexity of heterogeneous cloud particles down to simple quantities such as mass or number concentrations. These schemes also have documented deficiencies related to their assumptions of hydrometeor categories (cloud, rain, ice, snow) and the corresponding parameterizations that are required to convert particles between these categories (Igel, 2019; Lamb et al., 2023; Morrison, van Lier-Walqui, et al., 2020). Moreover, the parameterizations of one microphysics scheme are not necessarily modular or transferable to another, as baked-in design choices such as the assumption of a gamma or exponential distribution are not uniform between methods. As a result, today's climate modeler must navigate an immense menu of microphysics parameterizations without an option to improve the representation of hydrometeors systematically by incrementally relaxing controlled approximations.

Recent attempts to overhaul the microphysics method menu highlight the need for flexibility in climate modeling. These efforts range from identifying the more optimal moments of a particle distribution (Igel, 2019; Kogan & Belochitski, 2012; Morrison et al., 2019) or calibrating the parameters of similar moment-based microphysics frameworks (Azimi et al., 2023; Bieli et al., 2022; Morrison, van Lier-Walqui, et al., 2020), to overhauling the way hydrometeor distributions and process rates are represented (de Jong et al., 2022; Fierce & McGraw, 2017; Lamb et al., 2023; Rodríguez Genó & Alfonso, 2022; Seifert & Rasp, 2020). Indeed, as computational power increases and cloud microphysics becomes a leading source of uncertainty in climate predictions (Morrison, van Lier-Walqui, et al., 2020), the expense of more detailed and accurate microphysics methods becomes justified. Unfortunately, closing a set of equations for many higher-order moments is an ill-posed mathematical problem (Igel, 2019; Morrison, van Lier-Walqui, et al., 2020), which has limited the development or deployment of truly flexible moment-based microphysics methods.

We introduce a modernized moment-based approach to representing cloud microphysics that promises exactly that: flexibility. We build on the idea of a spectral approximation for the particle-size distribution (PSD) that is composed of many subdistributions (de Jong et al., 2022) by using a moment-based solution method that allows all parameters of the subdistributions to vary in time. This novel method presents a moving-abscissa version of the representation of de Jong et al., 2023, using a method of moments algorithm rather than the collocation numerical approach. Unlike the arbitrary-moment-predictor of Igel, 2019, our method evolves the microphysical governing equations directly in the space of moments, and unlike the generalized moment representation of Morrison, van Lier-Walqui, et al., 2020, we close the system of moments by assuming that subpopulations of particles follow physically interpretable distributions. Inspired by moving-bin or Lagrangian particle-based microphysics, this approach avoids mathematical ill-posedness by inverting the moments and parameters of sub-populations of the total PSD, much as a single Lagrangian superdroplet tracks the properties and multiplicity of a subpopulation of identical hydrometeors. In addition, our method handles particle-particle interactions (such as collisional coalescence) similarly to the superdroplet method (SDM) (Shima et al., 2009) by directly using pairwise particle process rates, rather than parameterized bulk conversion rates to transfer moments between subdistributions. This approach operates on a consistent and interpretable set of physical processes in the domain of particle properties. That is, rather than requiring scheme-specific parameterizations, the same process rates are used regardless of the number of moments and subdistributions tracked. The user can therefore explore different structural complexity by using more moments/subdistributions with the same microphysical process rates. These core process rates can even be calibrated in one model configuration (e.g. as in Bieli et al., 2022) and remain applicable to a different configuration of the method.

This work presents the numerical framework and results of this flexible moment-based microphysics method to represent warm-rain processes. We extend ‘Cloudy.jl‘ (Bieli et al., 2022) and other Julia packages from the Climate Modeling Alliance environment (‘KinematicDriver.jl‘) to implement and demonstrate the behavior and accuracy of this approach with reference to other standard moment-based schemes, as well as with reference to results of the Lagrangian SDM using the open-source implementation ‘PySDM‘. Section 2 presents the mathematical foundation for this method, followed by an evaluation of individual warm-rain process rates in section 3. Section 4 then brings these microphysics processes together with a simple advection scheme to simulate the formation and precipitation of an idealized one-dimensional cloud (Shipway & Hill, 2012) in comparison to standard bulk approaches. Finally, we offer additional insight and concluding remarks in section 5.

4.2 Method Description

4.2.1 Conceptual

Our proposed method can be conceptually described as a hybrid approach between Lagrangian microphysics methods and closure-based bulk microphysics (see Figure 4.1). In Lagrangian methods such as the superdroplet method (Shima et al., 2009), the PSD is described as a sum of pointwise or monodisperse PSDs, known as "superdroplets". Each superdroplet has properties including an associated mass and multiplicity (or number concentration), both of which can vary in time according to both single-particle processes such as condensation/evaporation, as well as from transfer of mass between superdroplets during particle-particle interactions such as collisional coalescence. A standard bulk approach tracks two categories of liquid hydrometeors (cloud and rain) through a commutable set of time-varying moments and a fixed-form subdistribution corresponding to each category (the "closure method of moments," or CMOM). In CMOM, b hyperparameters of the subdistribution are inferred from the first b moments. For instance, the hyperparameters of an exponential distribution include N (the number density) and θ (the particle mass scale), and are inferred from the first two moments in mass, M_0 and M_1, as $N = M_0$ and $\theta = M_1/M_0$. (A similar computation can be performed using three moments to infer the three parameters of a gamma or lognormal distribution.) Single-particle process rates are computed by integration over the size distribution implied by the current moments, whereas particle-particle process rates are typically specified as conversion rates which are fixed empirical functions of the moments (Khairoutdinov and Kogan, 2000 for instance).

Our proposed method maintains the CMOM with two key modifications: (1) we do not limit or require a particular number of subdistributions (such as cloud or rain); and accordingly, (2) transfer of mass (and all other moments) between subdistributions during particle-particle interactions is numerically computed or approximated on-the-fly, rather than relying on structure-specific conversion rates. Specifically, when considering the transfer of mass between a particle drawn from subdistributions of smaller and larger particles, the mass of the resulting coalesced hydrometeor (along with higher-order moments) is always transferred to the larger subdistribution, analogous to the bulk process of accretion (see Figure 4.2). When coalescence occurs between two particles drawn from the same subdistribution, the mass and moments of the resulting coalesced hydrometeor may either remain within this subdistribution or be transferred to a larger subdistribution. This decision is analogous to the difference between self-collection and autoconversion. However, rather than

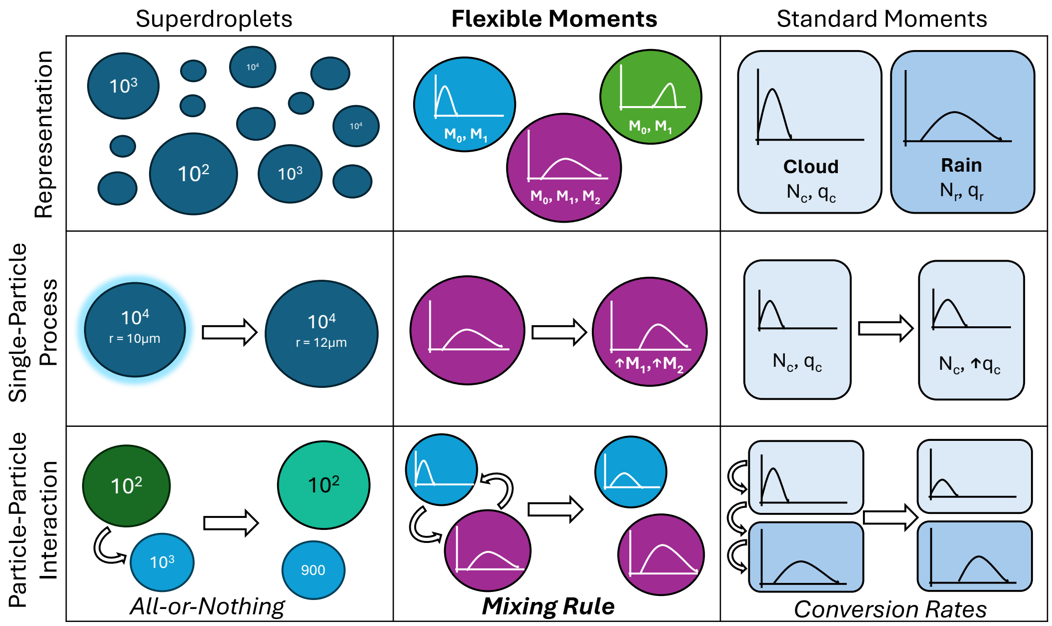


Figure 4.1: Conceptual illustration of the difference between the proposed implementation and the superdroplet method or a standard 2-moment microphysics scheme.

using integrated conversion rates, our computation utilizes a mixing rule/weighting function $W_k(x)$, which determines the fraction of each moment from a subdistribution k that remains in that subdistribution upon coalescence, depending on the particle size x and the other parameters of that subdistribution. For instance, the all-or-nothing mixing rule

$$W_k(x) = \begin{cases} 1 & \text{if } N_k > N_{k+1}, \\ 0 & \text{otherwise,} \end{cases}$$

partitions all mass to the subdistribution with larger number concentration as in the superdroplet method (Shima et al., 2009). Other forms such as a Heaviside function are considered later.

This reframing of standard 2-moment bulk microphysics allows us to extend this framework to a semi-Lagrangian style. An arbitrary number of subdistributions (and corresponding moments) may be used to describe a hydrometeor population while retaining the same set of parameterized process rates as particle-based microphysics, such as a rate of collisional-coalescence. In particular, the added flexibility of numerically computing particle-particle interactions eliminates the necessity of conversion rates that plague standard bulk microphysics, at relatively low computational cost.

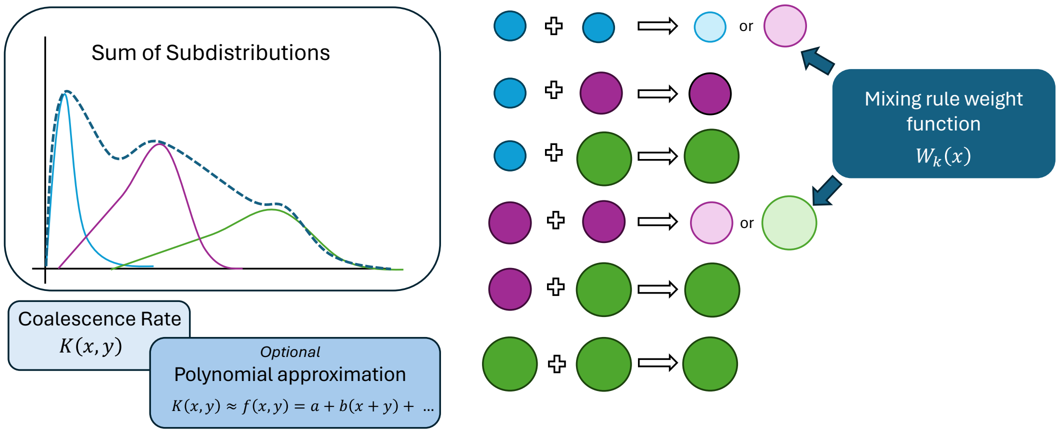


Figure 4.2: Illustration of transfer of mass and other moments between subdistributions according to a mixing rule.

The result is a flexible microphysics method which bridges the mathematical theory and complexity of Lagrangian and bulk microphysics methods, accessing a regime of high-fidelity microphysics which remains practical for large-scale global atmospheric simulations, unlike existing bin or superdroplet methods.

4.2.2 Mathematical

This work extends the frameworks of de Jong et al., 2022 and Bieli et al., 2022 to a flexible and arbitrary-moment representation based on the closure method of moments. Within an atmospheric gridbox, the particle size distribution $n(x)$ (where x is the particle mass) is approximated by a sum of several subdistributions $n_k(x)$,

$$n(x) = n_1(x) + n_2(x) + n_3(x) + \ldots, \tag{4.1}$$

where each subdistribution $n_k(x)$ may have a unique functional form. Further, each subdistribution has time-varying hyperparameters $\{N, \theta_1, \theta_2, \ldots\}$ such as the number density and mean mass for an exponential distribution, or the number density, scale factor, and shape factor for a gamma distribution. As in Bieli et al., 2022, there exists a well-defined map and inverse map between the hyperparameters and the first p moments of the subdistribution $\{M_{0,k}, \ldots, M_{p,k}\}$, where p is the number of hyperparameters. Then, at every point in time, the set of subdistribution moments uniquely defines the parameters of each subdistribution and thus the overall PSD. By contrast, inverting additional and higher-order overall moments of a multimodal distribution such as a gamma mixture does not necessarily lead to a stable or well-posed closure.

4.2.2.1 Single-particle processes

For a single-particle microphysical process rate $F_{sp}(x)$, which may represent quantities such as condensation/evaporation rates or sedimentation velocity and depends on particle size x, contributions of the process to each subdistribution and its moments can be computed independently because of the linearity of the composition (4.1):

$$\frac{d}{dt}M_j\bigg|_{sp} = \sum_k \frac{d}{dt}M_{j,k}\bigg|_{sp}, \tag{4.2}$$

where

$$\frac{d}{dt}M_{j,k}\bigg|_{sp} = \int_0^\infty x^j F_{sp}(x) n_k(x) dx. \tag{4.3}$$

In instances where $F_{sp}(x)$ can be well-approximated by a power series, these individual source terms are analytical sums of subdistribution moments. For instance, we consider a power series parameterization of the particle fall speed v_{fall} that allows for analytical computation of the moment-weighted fall speeds:

$$v_{\text{fall}}(x) = \sum_j v_j x^{\alpha_j}, \tag{4.4}$$

where v_j and α_j are empirically-derived parameters.

Furthermore, the processes of condensation/evaporation follow a power series analytical equation. Neglecting ventilation effects, the rate of change of a particle of radius r in an environment with saturation S can be described as

$$\frac{dr}{dt}\bigg|_{c/e} = \frac{S-1}{F_d + F_k}\frac{1}{r}, \tag{4.5}$$

where F_d and F_k are mass transfer coefficients (Pruppacher & Klett, 2010). Then the rate of change of the particle size distribution in terms of volume x is found to be

$$\frac{\partial r}{\partial t}\bigg|_{c/e} = -\frac{\partial}{\partial x}\Big(4\pi r n(x)(S-1)G(T)\Big), \tag{4.6}$$

where $G(T)$ encompasses the mass transfer coefficients and is assumed to be a function of temperature T. Finally, substituting for $r = \left(\frac{3}{4\pi\rho_l}x\right)^{1/3}$ with ρ_l referring to liquid water density,

$$\frac{\partial r}{\partial t}\bigg|_{c/e} n(x) = -\frac{(4\pi)^{2/3}3^{1/3}}{\rho_l^{1/3}}(S-1)G(T)\frac{\partial}{\partial x}\left(x^{1/3}n(x)\right). \tag{4.7}$$

Then, using the product rule and then performing integration by parts to find the kth moment, we obtain

$$\left.\frac{dM_k}{dt}\right|_{c/e} = 3\left(\frac{4\pi}{3}\right)^{2/3} \frac{1}{\rho_l^{1/3}}(S-1)G(T)kM_{k-2/3}. \tag{4.8}$$

In equations 4.8, the saturation S and temperature T are grid-mean quantities that are explicitly tracked by a fluid solver, whereas ρ_l is a fixed liquid density and $G(T)$ is an empirical function of temperature.

4.2.2.2 Collisional-coalescence

The collisional coalescence process is described by the stochastic collection equation,

$$\left.\frac{\partial n(x,t)}{\partial t}\right|_{cc} = \frac{1}{2}\int_0^x K(x-y,y)n(x-y)n(y)dy - n(x)\int_0^\infty K(x,y)n(y)dy, \tag{4.9}$$

where $K(x,y)$ is an empirical coalescence kernel that describes the rate of collisional coalescence between two particles of size x and y. The rate of change of the total PSD moment of order b is therefore

$$\frac{d}{dt}M^{(b)} = \frac{1}{2}\int_0^\infty x^b \int_0^x K(x-y,y)n(x-y)n(y)dydx - \int_0^\infty x^b n(x)\int_0^\infty K(x,y)n(y)dy\,dx. \tag{4.10}$$

In order to substitute our approximation of the PSD as a sum of subdistributions (equation (4.1)), our approach divides equation (4.9) into four component integrals:

- $Q_{j,k}^{(b)}$: rate of gain of moment order b in subdistribution k through collision with a particle distribution of smaller mean j (i.e., $k > j$). This term is akin to the increase in rain moments through accretion of cloud droplets:

$$Q_{j,k}^{(b)} = \frac{1}{2}\int_0^\infty x^b \int_0^x K(x-y,y)\Big[n_j(x)n_k(x-y) + n_k(x)n_j(x-y)\Big]dy\,dx. \tag{4.11}$$

- $R_{j,k}^{(b)}$: rate of loss of moment order b from subdistribution j through collision with subdistribution k, with $j \neq k$. Note that this term is symmetric with Q when $b = 1$ (implying mass conservation):

$$R_{j,k}^{(b)} = -\int_0^\infty x^b n_k(x)\int_0^\infty K(x,y)n_j(y)dy\,dx. \tag{4.12}$$

- $S_{+,k}^{(b)}$: rate of change of moment order b due to internal collisions within subdistribution k that remain internal to subdistribution k (akin to self-collection), defined according to a weighting function (i.e. mixing rule) $W_k(x)$:

$$S_{+,k}^{(b)} = \frac{1}{2}\int_0^\infty x^b \int_0^x K(x-y,y)n_k(x)n_k(x-y)W_k(x)dy\,dx. \tag{4.13}$$

- $S_{-,k}^{(b)}$: rate of change of moment order b due to internal collisions within particle k that are transferred to particle $k+1$ (akin to autoconversion), defined using the same weighting function $W_k(x)$. Note that $S_{-,P}^{(b)} = 0$ where P is the total number of subdistributions:

$$S_{-,k}^{(b)} = \frac{1}{2}\int_0^\infty x^b \int_0^x K(x-y,y)n_k(x)n_k(x-y)\left[1 - W_k(x)\right]dy\,dx. \tag{4.14}$$

The moments of each subdistribution k then satisfy the evolution equation

$$\frac{d}{dt}M_k^{(b)} = \sum_{j=1}^{k-1} Q_{j,k}^{(b)} + \sum_{j=1}^{P} R_{j,k}^{(b)} + S_{+,k}^{(b)} + S_{-,k-1}^{(b)} \tag{4.15}$$

with $S_{-,0}^{(b)} = 0$. It can be verified that the sum of these subdistribution moment tendencies satisfy Equation (4.10) for the overall PSD moments.

The innovation of this technique of separating integrals is twofold. First, the integrals $Q_{j,k}^{(b)}$ and $R_{j,k}^{(b)}$ can be solved analytically for most distribution closures (such as a gamma or exponential) when the kernel $K(x,y)$ is expressed as a power series or polynomial approximation. (See Bieli et al., 2022 for the full derivation.) Second, the use of a weighting function $W(x)$ in the self-collection and autoconversion integrals $S_{+,k}^{(b)}$ and $S_{-,k-1}^{(b)}$ allows for generalization of the method to more than two categories without the need for additional conversion rates while maintaining physical consistency through numerical integration over the collision kernel $K(x,y)$.

The weighting function describes the fraction of collisions within a subdistribution that act as moment source terms to the same subdistribution (self-collection) versus source terms to the next-largest subdistribution (autoconversion). For instance, a commonly used coalescence algorithm in the superdroplet method uses an all-or-nothing weighting depending on the superdroplet multiplicity. An alternative continuous choice might be the unweighted marginal density of subdistribution k for a particle size x, i.e.

$$W_k^{(b)}(x) = \frac{n_k(x)/N_k}{\sum_{j=1}^{L} n_j(x)/N_j}, \tag{4.16}$$

where L is the total number of subdistributions and N_j is the number density parameter of subdistribution j. A physically-motivated choice for the weighting function might use a Heaviside function with fixed size thresholds x_k^*,

$$W_k^{(b)}(x) = 1 - \mathcal{H}(x - x_k^*), \tag{4.17}$$

such that the domain of subdistribution k is fixed to $[x_{k-1}^*, x_k^*]$. This choice effectively transforms the subdistribution closure functions to finite elements in the particle size domain, as in a related spectral method (de Jong et al., 2022). It is further motivated by the tradition of thresholding the size of cloud droplets to a value such as 25 μm or 40 μm (Igel et al., 2022) above which rates of self-collection and accretion rapidly accelerate. Finally, the use of mass thresholding reduces the numerical integration required by Equations 4.13-4.14 to a one-dimensional integral, rather than a more computationally expensive two-dimensional integration. Because of these computational advantages and connection to standard presentations of cloud and rain content in microphysical simulations, our demonstrations will use this Heaviside form of the weight function.

4.2.2.3 Advection

Advection of moments between gridboxes in the flexible method can proceed in an Eulerian, Lagrangian, or hybrid approach, depending on the context including gridbox spacing and number of subdistributions employed.

1. **Eulerian** Similarly to traditional bulk microphysics, every gridbox maintains a set of corresponding subdistributions and moments. Advection proceeds by transferring moments to their corresponding tracers in neighboring gridbox, and the advection scheme may employ additional subgrid-scale transport closures. This technique has the advantage of consistent tracers within gridboxes, but a potential disadvantage of introducing moments to spatially adjacent subdistributions with vastly different properties, for instance at the edge of a cloud.

2. **Lagrangian** Akin to the superdroplet method, subdistributions (and associated moments) retain a notion of position in the overall computational domain and are advected by the mean flow. Subdistributions are then ordered by mean particle size within each grid box in order to compute the coalescence integrations (Equations 4.11-4.15). This method may be performant when a large number of subdistributions are employed.

3. **Hybrid** One mechanism to alleviate potential subdistribution mis-match in the Eulerian advection approach is by transferring advected moments to the most-similar subdistribution in the adjacent gridbox, rather than the subdistribution with the same index. Although we do not test this approach in this work, we suggest this hybrid technique for future implementations.

For simplicity, demonstrations in this work utilize the Eulerian advection approach. We leave detailed evaluation of alternatives to future work.

4.2.3 A simple example: collisions in a box

We demonstrate the behavior of this numerical method in a single gridbox (zero-dimensions) as it transfers mass between subdistributions during collisional coalescence in Figure 4.3. This example approximately solves equation 4.9 (the stochastic collection equation) using the mathematical procedure described by equations 4.10–4.15. Figure 4.3 illustrates the evolution of the spectra, first moment, and subdistribution parameters in the case of a Golovin or linear coalescence kernel of the form $K(x, y) = B(x + y)$, with $B = 5$ m^3 kg^{-1} s^{-1} and with x and y expressed in kg. The initial distribution is specified as an exponential distribution with $N = 10^8$ m^{-3} and scale $\theta = 10^{-10}$ kg, and four gamma modes are used to represent the evolution of a total of 12 moments. We use a Heaviside function as the weighting term for internal collisions with mass thresholds of 10^{-9}, 10^{-7}, and 10^{-5} kg for the first three modes. As the system evolves, mass is transferred from smaller modes to larger ones. Larger and larger modes are activated as particles collide, coalesce, and their moments are transferred to subsequently larger subdistributions. For instance, the second subdistribution (orange) initially experiences a growth in its first moment due to autoconversion from the first subdistribution (blue), and subsequently accretion of this smaller mode. Its mass scale correspondingly increases as the average particle size represented by this mode increases, until reaching the threshold determined by the weighting function, and the shape factor evolves to capture the broadening and then narrowing of the subdistribution. Selecting different mass thresholds for the Heaviside weight function (such as uneven spacing) impacts the moments and parameters of individual subdistributions, but does not strongly affect the overall PSD or moments as the system evolves in time in the case of this simple linear kernel. However, for more complex kernels which behave differently in different particle size regimes (such as a geometric kernel), setting the mass thresholds to correspond with these regimes is advantageous.

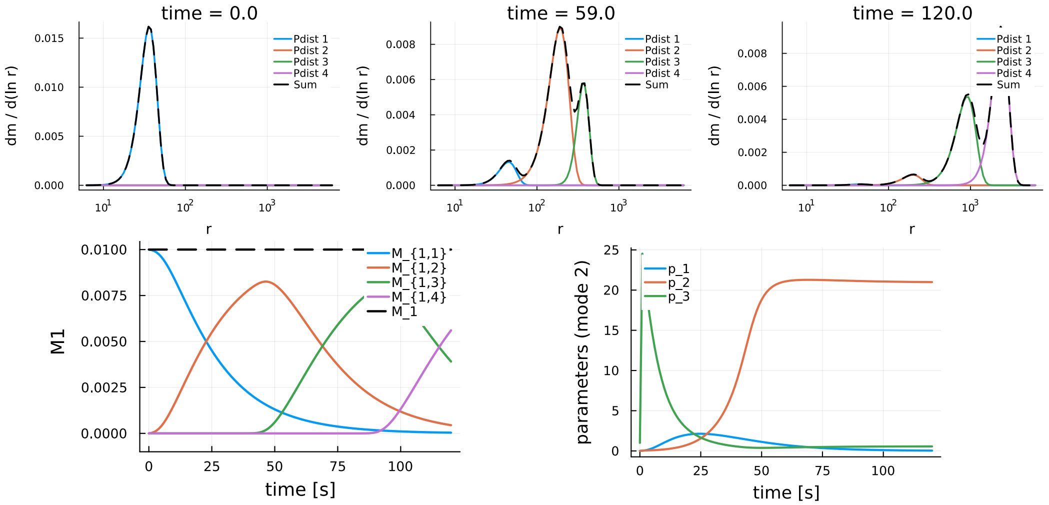


Figure 4.3: Evolution of the particle size distribution (top) under a linear coalescence kernel using four gamma modes in the proposed method. Also shown are the evolution of the first moment (mass density) within each of the subdistributions (bottom left), and the evolution of the parameters of the second gamma subdistribution (bottom right; shown are number density cm^{-3} (blue), mass scale in μg (orange), and shape factor (green)).

4.3 Evaluation of subgrid scale hydrometeor processes

We now evaluate the performance of the proposed method in representing the subgrid-scale processes of coalescence and condensation in a zero-dimensional setting. For reference, we include results from the superdroplet method implementation PySDM (Bartman et al., 2022; de Jong et al., 2023), or relevant literature results. Although Lagrangian methods have been shown to display substantial disagreement in representing rain formation due to differences in coalescence and initialisation (Hill et al., 2023; Unterstrasser et al., 2017), the results are included here as a one-to-one comparison using the same kernel and terminal velocity parameterizations as the flexible method with a large number of superdroplets. The succeeding section will combine subgrid-scale dynamics with vertical advection to illustrate the formation and precipitation of a one-dimensional warm cloud, again with reference to a particle-based simulation.

4.3.1 Collisions in a box

First we consider two examples of zero-dimensional droplet coalescence, again solving the stochastic collection equation (Eq. 4.9) using the flexible method of moments. In Figure 4.4 we illustrate the moments and spectra obtained by using the

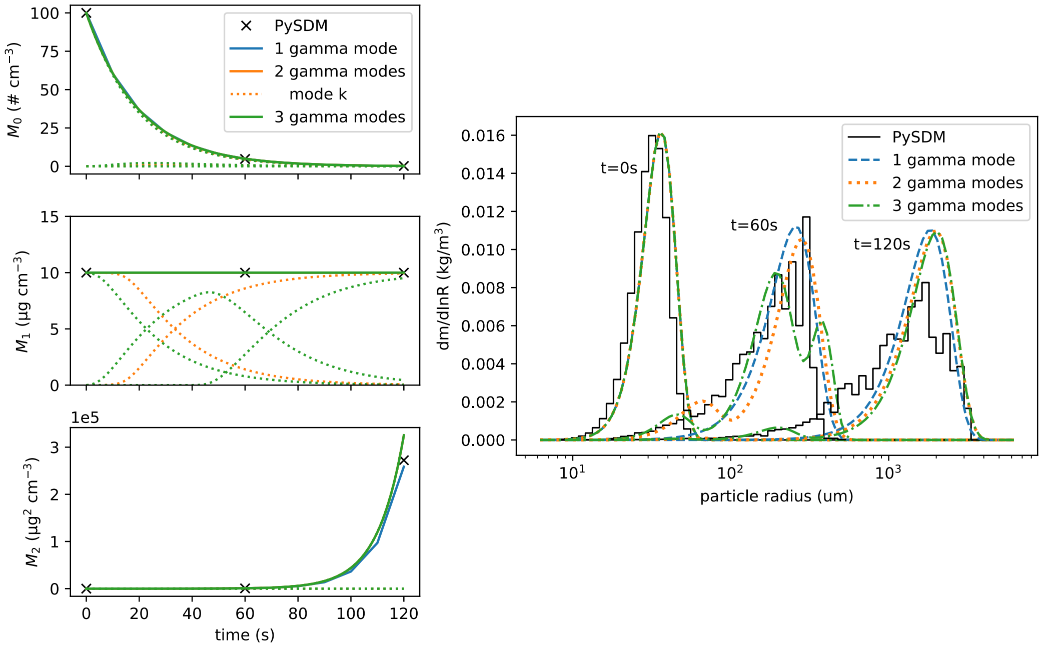


Figure 4.4: Evolution of the first three moments (left) and snapshots of the PSD (right) for a particle distribution evolving under a linear coalescence kernel, solved with either PySDM (superdroplets), a single gamma method-of-moments, or the proposed method with two or three gamma subdistributions.

same linear Golovin kernel dynamics and initial condition as in Figure 4.3, solved using a flexible-moment system in Cloudy.jl with either one, two, or three gamma subdistributions, corresponding to 3, 6, or 9 moments. In the case of these linear dynamics, the global-domain integrals Q and R are computed analytically, and the mass-transfer integrals $S_{+/-}$ are computed semi-analytically for the multi-mode configurations with a Heaviside weighting function and mass threshold $x^* = 4\mu g$ (approximately 100μm radius) in the case of two modes, and 1μg and 100μg (62μm and 287μm radii) for three modes. For comparison, we include results from the stochastic superdroplet method using 2^{12} superdroplets and a maximum adaptive time step of 1s (Bartman et al., 2022). Both the spectra and moments appear the same as the SDM results as particle sizes evolve from 10's to 1000's of μm size. There is almost no difference between the 1-mode and 2- or 3-mode versions of our proposed method, except for the appearance of a tiny smaller mode at 60s simulation time that implies a broader intermediate spectrum. The evolution of the first moment shows that indeed by 60s, most mass has been transferred from smaller to larger modes in the 2- and 3-mode instances.

Next we consider a more difficult parameterization of the collisional kernel: a geometric coalescence kernel of the form $K(x, y) = C\pi(r(x)^2+r(y)^2)\big|A(x)-A(y)\big|$ where $r(x)$ is the radius and $A(x)$ is the cross-sectional area for a droplet of mass x, with $C = 100\text{m}^3\text{kg}^{-4/3}\text{s}^{-1}$. We approximate this non-polynomial kernel as an order 4 polynomial (see figure 4.10) in order to perform the integrations of Q, R, and S analytically or semi-analytically. Figure 4.5 includes an additional flexible moment configuration using 3 exponential modes to compare the utility of tracking moments of more or less complex distributions. The superdroplet reference simulations use 2^15 superdroplets and a 0.5s adaptive time step. From Figure 4.5 we see that this nonpolynomial kernel is more challenging to capture with only 6 moments, but also that there is a clear benefit to including more subdistributions. Although the 2-mode gamma closure and the 3-mode exponential closures both utilize 6 moments, the 3-mode version is better able to capture the broadening of the PSD over time, as well diagnosing the super-exponential growth in the second moment despite not tracking the second moments prognostically. Unlike the linear kernel in Figure 4.4, where the PSD follows an approximately gamma distribution at all times, these closure assumptions are a poor approximation for the broad distribution resulting from the hydrodyamic kernel. Indeed, use of a larger quantity of simpler exponential subdistributions helps to alleviate this closure error by improving the representation of persistent smaller (sub-100μm) particles, which collide at a slower rate than the larger droplets. This 3-mode exponential configuration in fact outperforms the 3-mode gamma configuration, suggesting that higher-complexity closures (i.e. higher moment-order methods) provide diminishing gains beyond two subdistributions.

4.3.2 Condensation in a parcel

The evolution of the moments of a PSD under condensational growth (derived analytically in equation 4.8) are demonstrated in an adiabatic parcel simulation following the setup of Rogers, 1975. The parcel begins unsaturated and rises adiabatically at a fixed velocity of 10ms^{-1}, condensing liquid water mass onto a fixed number concentration of $200cm^{-3}$ droplets. Using a monodisperse or single gamma distribution in our method matches the original parcel model results of Rogers, 1975 quite closely, with saturation leveling off near 1% as the rate of condensation balances increased relative humidity from the rising parcel. An additional test case demonstrating the evolution of a 2-mode mixture of an exponential and gamma distribution demonstrates the impact of the size distribution shape on the maximum attained saturation. Even though this distribution is initialized with the same total mass and number concentration as the other cases, the distribution is skewed toward

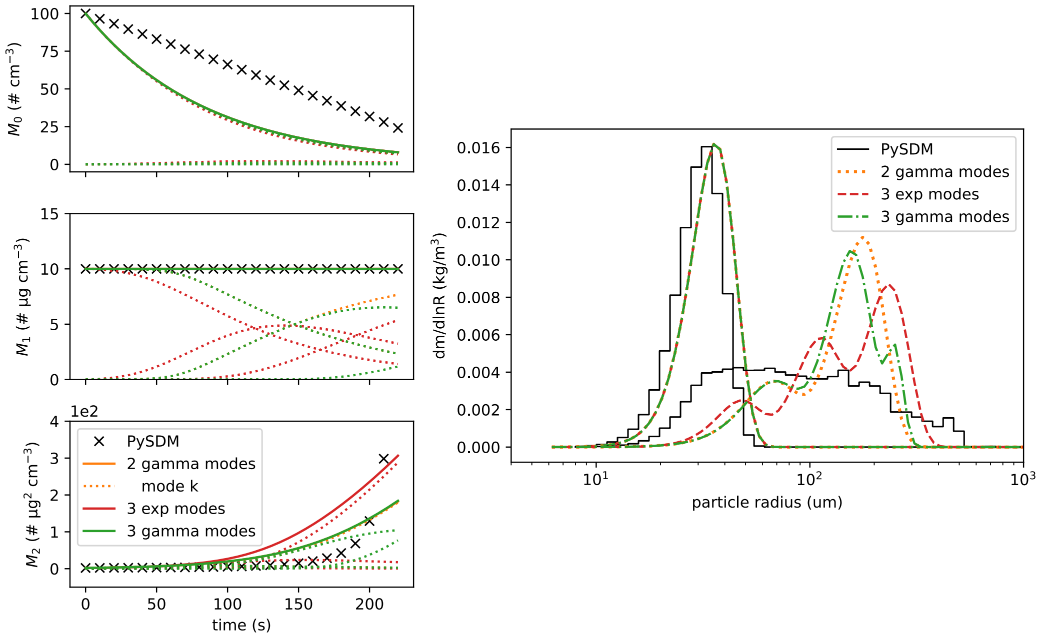


Figure 4.5: Evolution of the first three moments (left) and snapshots of the PSD (right) for a particle distribution evolving under a geometric coalescence kernel, solved with either PySDM (superdroplets), or the proposed method with two or three gamma subdistributions or with three exponential subdistributions.

larger hydrometeors, resulting in slower condensational growth and therefore less depletion of the supersaturation field. These results demonstrate the capacity of this unified and flexible method-of-moments framework to capture additional subgrid-scale process detail that is unresolved by standard saturation-adjustment calculations.

4.4 Comparison in a One-Dimensional Rainshaft

4.4.1 Comparison with Lagrangian Microphysics

Finally we combine the subgrid scale dynamic of coalescence (Equation 4.10) with vertical advection (including computation of moment-weighted terminal velocities, Equation 4.4) to simulate a one-dimensional precipitating cloud by solving for subdistribution moments as a function of time and altitude. Figure 4.7 demonstrates the vertical profiles of cloud and rain number (N_c, N_r respectively) and mass concentrations (q_c and q_r) as the cloud evolves in time using two gamma subdistributions (6 moments) in the flexible method, again with reference to Lagrangian microphysics simulations from PySDM. The simulations are initialised to contain liquid cloud water following a gamma size distribution between 1.5km and 2.25km altitude, with

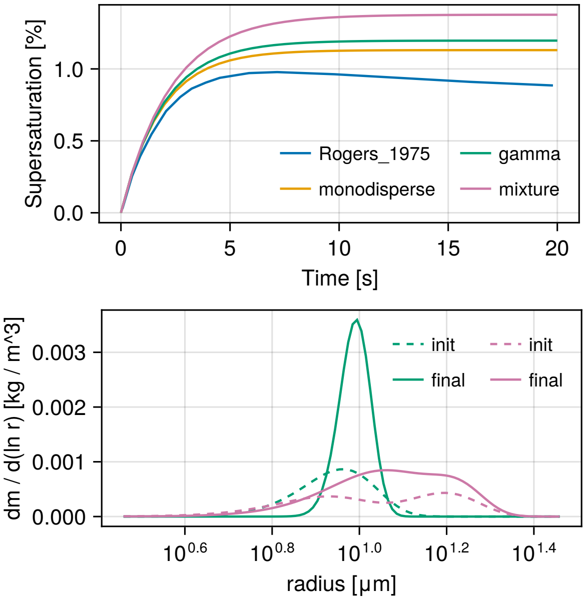


Figure 4.6: Adiabatic parcel simulation based on the setup of Rogers, 1975 with three different initial PSDs: monodisperse, gamma distribution, and mixture of an exponential and gamma distribution.

no liquid water elsewhere. Both the flexible moment method and the SDM reference simulations use a 1s timestep and 150m vertical grid spacing. Both methods use a power-law terminal velocity of $v_t(x) = V_0 x^{\beta}$ with $V_0 = 50\text{ms}^{-1}\text{kg}^{-1/6}$ and $\beta = 1/6$, as well as a Golovin coalescence kernel with $B = 5\text{m}^3\text{kg}^{-1}\text{s}^{-1}$ as in previous examples. For visualization purposes, we distinguish between "cloud" and "rain" using a fixed size threshold of 50μm radius. This distinction results in some initial rain concentration and mass appearing in both profiles, as the initial particle size distribution is a globally-supported gamma distribution.

The profiles in Figure 4.7 show a small underestimate in conversion of cloud number concentration to rain by the flexible method in the first 250s of simulation time, while conversion of mass concentration is initially overestimated, reflecting a small underprediction of the left tail (quantity of small particles) of the PSD in Figure 4.4. Advection of the Eulerian moments according to a first-order upwinding scheme results in vertical broadening of the cloud over time relative to the Lagrangian SDM. At later times these mass and number concentrations more closely match, as does the altitude of maximum rain mass concentration as rain sediments. Surface precipitation quantity and timing are well predicted by our method. Overall these results confirm that our method is able to reproduce a particle-based simulation

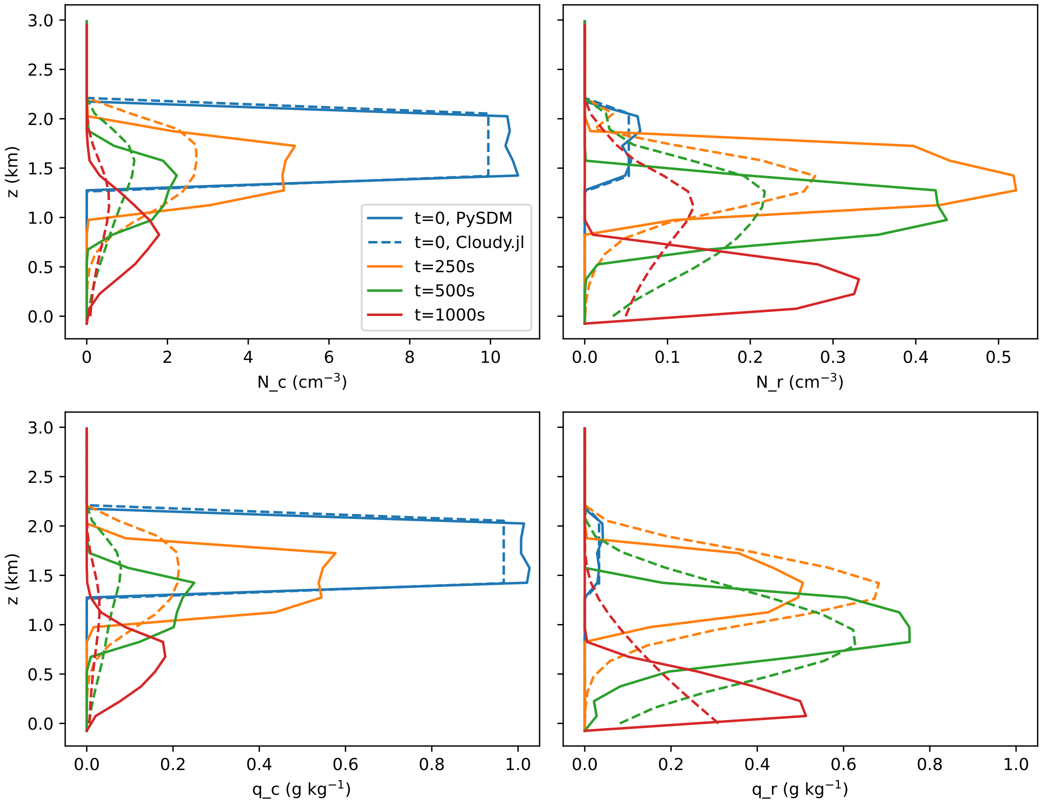


Figure 4.7: 1-dimensional rainshaft results comparing the proposed method (with coalescence and sedimentation) against Lagrangian microphysics with the same process rates. Subpanels are, clockwise from top left: cloud number concentration, rain number concentration, cloud specific humidity, and rain specific humidity.

given the same definition of microphysical processes, unlike tradition bulk schemes whose structure-specific parameters would need to be re-calibrated (Azimi et al., 2023).

4.4.2 Comparison with Standard Bulk Methods

Finally, we illustrate the results of similar 1D cloud with coalescence and sedimentation processes in reference to other standard bulk microphysics schemes and Lagrangian microphysics. Figure 4.8 illustrates the time-altitude evolution of a cloud where small cloud droplets have been initialized uniformly (and consistently between simulations) throughout the domain, then left to coalescence and precipitate according to the microphysics scheme and parameterizations. We contrast the differences within the flexible moment framework that arise from changing either the model structure (the number of moments) or the collision rate parameterization, with comparison to the result of changing between traditional bulk microphysics representations and parameterizations.

In Figure 4.8 it becomes apparent that the parameterizations used within a standard bulk moment scheme can have a larger impact on the production and timing of precipitation than the structure of the model (1M versus 2M). The empirical 1-moment parameterizations of KK2000 (which is stractocumulus-specific) (Khairoutdinov & Kogan, 2000) and Kessler (Kessler, 1969) are more different from each other than from a completely different 2-moment parameterization. This 2-moment SB2006 (Seifert & Beheng, 2006) example is derived from the coalescence parameterization of Pinsky et al., 2001, which has a similar piecewise structure as the Long kernel (Long, 1974), but with limiters on rain formation. In contrast, substituting the collision kernel within our flexible moment-based approaching or using only 4 moments rather 6 yields results which are much more self-consistent. Using Long's kernel results in only a slight delay and reduction in precipitation relative to the linear kernel, owing to the slower initial conversion of cloud to rain in the linear small-particle size range. Likewise, using 4 moments rather than 6 (i.e. two exponential distributions, rather than two gamma distributions) slightly reduces and delays precipitation due to a less accurate representation of the PSD tail by the simpler exponential closure. As seen in figure 4.7, when using the same Golovin collision kernel, our method closely matches Lagrangian particle-based results (g). These results exemplify both the challenges inherent to current microphysics modeling practices and the power of using a flexible moment-based tool that does not require structure-specific conversion parameterizations. The modeler need only specify the collision kernel itself, be it linear, Long's, or a more complex kernel such as Berry 1967 (Berry, 1967) without regard for the number of moments or subdistributions used to represent the PSD.

In figure 4.9 we include the process of cloud formation through aerosol activation (Abdul-Razzak & Ghan, 2000) and condensation following the kinematic framework of Shipway and Hill, 2012 to further highlight the difference between the use of bulk conversion rates or collision kernels in the flexible and Lagrangian microphysics methods. In these simulations, the fluid field is decoupled from the microphysical processes rates, and only condensation/evaporation feeds back on the local saturation field. As in Seifert and Beheng, 2006, we impose tendency limiters on the number concentration in the flexible microphysics method to prevent artificial evaporation that would lead to artificially small (and slowly sedimenting) particles. The cloud initially forms at the same time and altitude in all three settings, as the

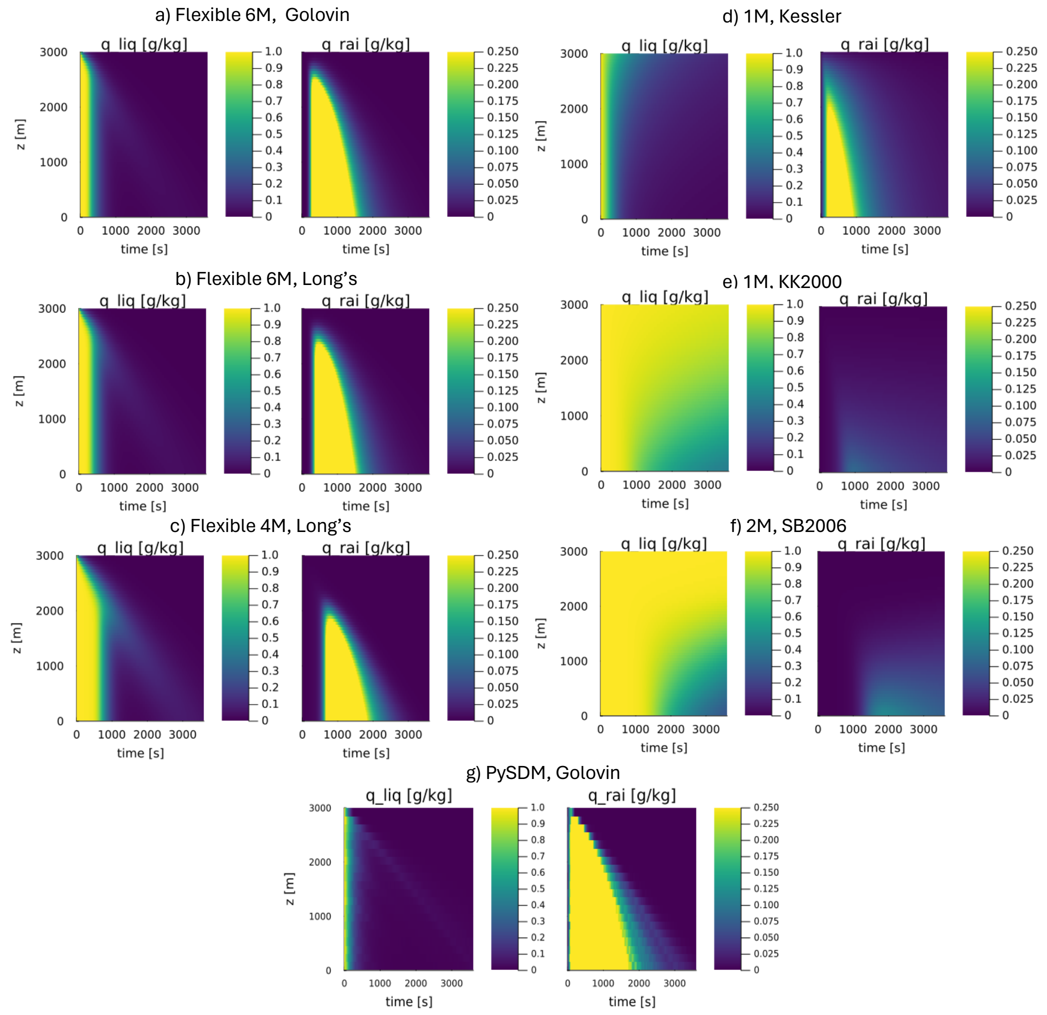
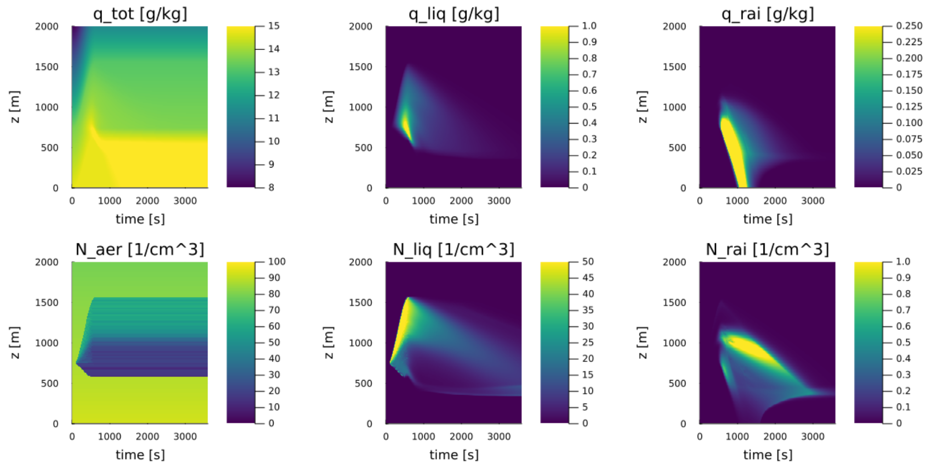
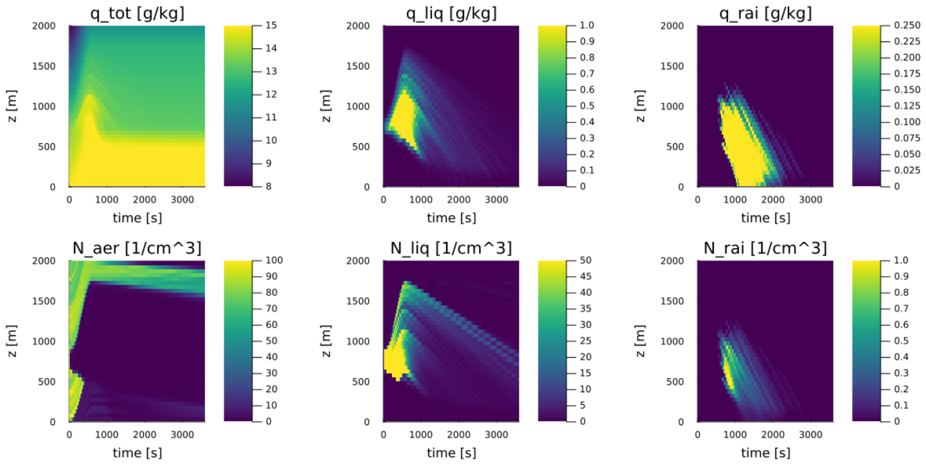


Figure 4.8: Liquid and rain water content for a coalescence-sedimentation version of the 1-dimensional rainshaft, solved using six different microphysics parameterizations. Left, from top: (a) the proposed flexible method of moments with two gamma modes (6 moments) and a linear collision kernel, (b) the same but with Long's collision kernel (Long, 1974), and (c) with two exponential modes (4 moments) and Long's kernel. Right, from top: (d) a single-moment (mass concentration) representation of cloud and rain using either the Kessler precipitation rate (Kessler, 1969) or (e) the KK2000 autoconversion parameterizations (Khairoutdinov & Kogan, 2000), (f) and a two-moment (mass and number concentration) representation using the parameterizations of SB2006 (Seifert & Beheng, 2006). Bottom: (g) Lagrangian particle-based results simulated in PySDM (Bartman et al., 2022; de Jong et al., 2023) with the same Golovin collision kernel and sedimentation scheme as in (a).

a) Flexible 6M, Golovin



b) PySDM, Golovin



c) 2M, SB2006

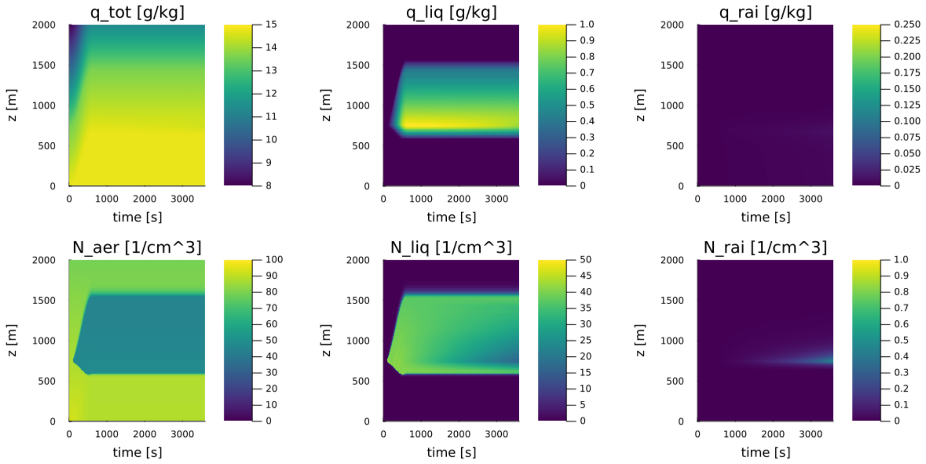


Figure 4.9: Water content and number concentrations for the full 1-dimensional idealized simulation framework of Shipway and Hill, 2012 using (a) the flexible moment framework with 6 moments and a Golovin kernel, (b) Lagrangian microphysics (PySDM) with the same Golovin kernel, and (c) the double-moment parameterization of Seifert and Beheng, 2006.

aerosol-activation parameterization is consistent. However, while cloud droplets grow through bulk thermodynamic equilibrium in the 2M scheme, condensation is explicitly resolved in the flexible and Lagrangian methods. The flexible method most notably shows much quicker onset of precipitation and depletion of cloud water, which are consistent with the Lagrangian results and attributable to differences in both the representation of coalescence and sedimentation rates. As seen in figure 4.7, the Eulerian advection of hydrometeors in the flexible method leads to vertical diffusion of the rain hydrometeors in comparison with the Lagrangian method. All liquid particles sediment at their respective terminal velocities in the flexible method (aerosols are not tracked explicitly, unlike in the Lagrangian SDM), whereas most bulk schemes only consider sedimentation of the rain category. The 2-moment SB2006 cloud is therefore much more persistent and does not show the sedimentation behavior of cloud particles in the flexible or Lagrangian methods. Indeed, the cloud number concentration is relatively uniform throughout the 2M cloud, indicating a more homogeneous droplet size distribution, and precipitation only appears near cloud base. This behavior is reflective of the built-in assumptions Seifert and Beheng, 2006 regarding the shape of the PSD for cloud and rain modes. This distinction further highlights the benefit of directly using consistent and generalized process rates such as coalescence kernel rates or terminal velocity across all modes or subdistributions, rather than partitioning processes between cloud and rain. The flexible moment framework therefore produces cloud simulations which are both mathematically consistent, and consistent with Lagrangian SDM results.

4.5 Discussion and Conclusions

This work presents a flexible moment-based microphysics method as a potential solution to the longstanding "warm rain problem", which refers to the inability of parameterized conversion rates to accurately describe rain formation in bulk microphysics. Our method approximates the particle size distribution as sum of an arbitrary number of modes (or subdistributions), each with associated moments, rather than using a fixed bimodal cloud-rain descriptions as in most traditional methods. We therefore rely on fundamental process rates such as particle-particle coalescence or condensation which are dependent on particle size, rather than empirical rates of conversion between cloud and rain categories which are not necessarily generalizable. We show that by using the same set of particle-based process rates, this method can reproduce Lagrangian superdroplet simulations in a box and one-dimensional setting. In fact, with an appropriate choice of the weighting function and subdistributions which are delta functions, this flexible method mathematically converges toward the expectation value of a particle-based representation when many subdistributions are used.

The flexible structure of this approach means that not only can the modeler use a wide range of complexity to represent hydrometeors (by specifying fewer or more subdistributions), but also that the microphysical process rates generalize to any such different configuration. This is in contrast to traditional bulk approaches, which often have structure-dependent assumptions and parameterizations of key processes that contribute to rain formation. As a result, this flexible approach to microphysics presents a useful research tool to decorrelate and explore the accuracy-complexity tradeoff of both the hydrometeor representation and the core process rates.

Finally, the similarities between this approach and Lagrangian microphysics may enable straightforward extensions of the method beyond warm rain microphysics. Analogously to a superparticle, a subdistribution and its corresponding of moments may be described with either fixed or covarying properties in the aerosol or ice space, such as dry aerosol size, hygroscopicity, ice density, or surface area. As a result, this flexible warm rain approach may be seamlessly integrated with existing parameterizations of mixed phase microphysics (by using fixed additional properties), or even used to unify the representation of additional traditional hydrometeor categories from the point of aerosol nucleation and coagulation, to the status of a precipitating rain droplet, hailstone, or snowflake.

4.6 Appendix A: Polynomial approximation of coalescence kernels

When $K(x, y)$ is a polynomial, then the SCE moment integrals Q and R can be computed analytically, improving the efficiency of our method. The approximation of a given kernel function as an order r polynomial is performed through a nonlinear optimization. The order-zero term in the polynomial series is defined as $K(0, 0)$, and subsequent higher order terms are the minimizer of an objective function defined as the residual L_2 error of the approximation evaluated on a 2D grid of candidate colliding particle masses. Two examples of this polynomial approximation are illustrated in figure 4.10.

Open Research

The implementation described in this paper can be found at https://github.com/CliMA/Cloudy.jl v0.5.3, with additional examples generated using https://github.com/CliMA/KinematicDriver.jl and PySDM https://github.com/open-atmos/PySDM.

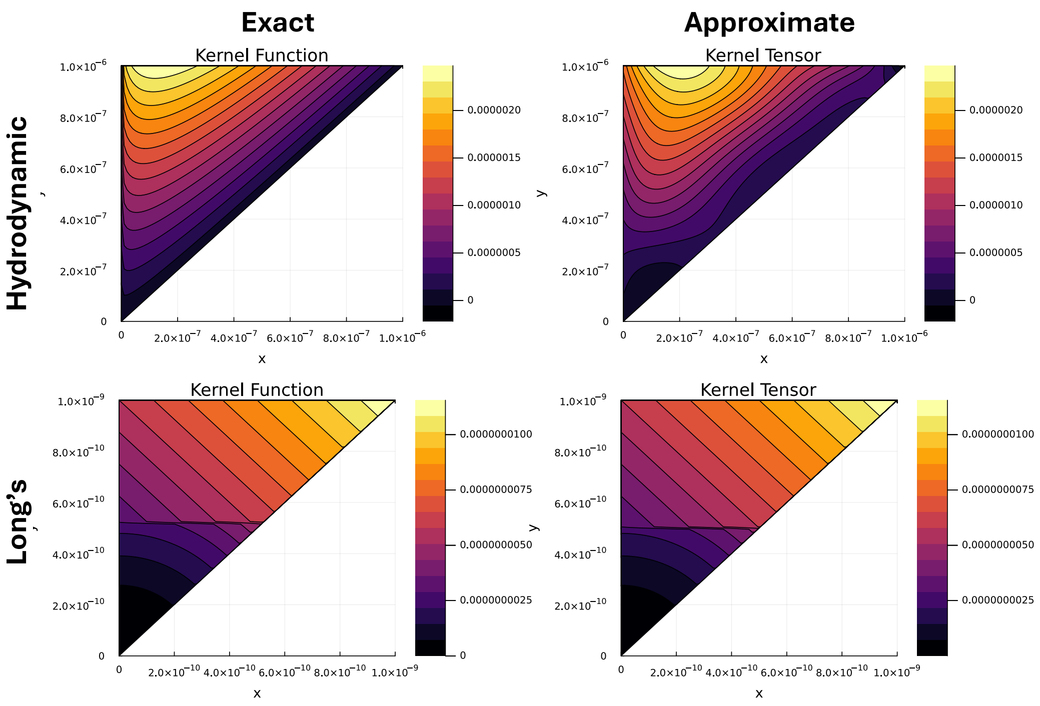


Figure 4.10: Illustration of the exact (left) and polynomial approximations (right) for two non-polynomial collision kernels as a function of the two colliding particle masses (in kg): a geometric kernel (order 4 approximation; top) and Long's kernel (Long, 1974) (order 2 approximation; bottom).

Acknowledgements

This work was coauthored by Sajjad Azimi, Anna Jaruga, and Tapio Schneider. E. de Jong was supported by a Department of Energy Computational Sciences Graduate Fellowship. S. Azimi acknowledges support from Swiss National Science Foundation (SNSF, 678 Grant No. P500PN 202876). This work was additionally supported by Schmidt Sciences, LLC. The authors are grateful to Nathaniel Efrat-Henrici, Melanie Bieli, Tobias Bischoff, and other researchers and developers of the Climate Modeling Alliance software environment for their contributions to ‘Cloudy.jl‘ and peripheral packages used in this work.

References

Abdul-Razzak, H., & Ghan, S. J. (2000). A parameterization of aerosol activation: 2. Multiple aerosol types. *Journal of Geophysical Research: Atmospheres*, *105*(D5), 6837–6844. https://doi.org/10.1029/1999JD901161

Azimi, S., Jaruga, A., de Jong, E. K., Arabas, S., & Schneider, T. (2023). Training warm-rain bulk microphysics schemes using super-droplet simulations. https://doi.org/10.22541/essoar.169603562.20867617/v1

Bartman, P., Bulenok, O., Górski, K., Jaruga, A., Łazarski, G., Olesik, M. A., Piasecki, B., Singer, C. E., Talar, A., & Arabas, S. (2022). PySDM v1: Particle-based cloud modeling package for warm-rain microphysics and aqueous chemistry. *Journal of Open Source Software*, *7*(72), 3219. https://doi.org/10.21105/joss.03219

Berry, E. X. (1967). Cloud Droplet Growth by Collection. *Journal of Atmospheric Sciences*, *24*(6), 688–701. https://doi.org/10.1175/1520-0469(1967)024<0688:CDGBC>2.0.CO;2

Bieli, M., Dunbar, O. R. A., de Jong, E. K., Jaruga, A., Schneider, T., & Bischoff, T. (2022). An Efficient Bayesian Approach to Learning Droplet Collision Kernels: Proof of Concept Using "Cloudy," a New n-Moment Bulk Microphysics Scheme. *Journal of Advances in Modeling Earth Systems*, *14*(8), e2022MS002994. https://doi.org/10.1029/2022MS002994

de Jong, E. K., Bischoff, T., Nadim, A., & Schneider, T. (2022). Spanning the Gap From Bulk to Bin: A Novel Spectral Microphysics Method. *Journal of Advances in Modeling Earth Systems*, *14*(11), e2022MS003186. https://doi.org/10.1029/2022MS003186

de Jong, E. K., Singer, C. E., Azimi, S., Bartman, P., Bulenok, O., Derlatka, K., Dula, I., Jaruga, A., Mackay, J. B., Ward, R. X., & Arabas, S. (2023). New developments in PySDM and PySDM-examples v2: Collisional breakup, immersion freezing, dry aerosol initialization, and adaptive time-stepping. *Journal of Open Source Software*, *8*(84), 4968. https://doi.org/10.21105/joss.04968

Fierce, L., & McGraw, R. L. (2017). Multivariate quadrature for representing cloud condensation nuclei activity of aerosol populations. *Journal of Geophysical Research: Atmospheres*, *122*(18), 9867–9878. https://doi.org/10.1002/2016JD026335

Hill, A. A., Lebo, Z. J., Andrejczuk, M., Arabas, S., Dziekan, P., Field, P., Gettelman, A., Hoffmann, F., Pawlowska, H., Onishi, R., & Vié, B. (2023). Toward a Numerical Benchmark for Warm Rain Processes. *Journal of the Atmospheric Sciences*, *80*(5), 1329–1359. https://doi.org/10.1175/JAS-D-21-0275.1

Igel, A. L., Morrison, H., Santos, S. P., & van Lier-Walqui, M. (2022). Limitations of Separate Cloud and Rain Categories in Parameterizing Collision-Coalescence for Bulk Microphysics Schemes. *Journal of Advances in Modeling Earth Systems*, *14*(6), e2022MS003039. https://doi.org/10.1029/2022MS003039

Igel, A. L. (2019). Using an Arbitrary Moment Predictor to Investigate the Optimal Choice of Prognostic Moments in Bulk Cloud Microphysics Schemes. *Journal of Advances in Modeling Earth Systems*, *11*(12).

Kessler, E. (1969). On the Distribution and Continuity of Water Substance in Atmospheric Circulations [doi: 10.1007/978-1-935704-36-2_1]. In E. Kessler (Ed.), *On the Distribution and Continuity of Water Substance in Atmospheric Circulations* (pp. 1–84). American Meteorological Society.

Khairoutdinov, M., & Kogan, Y. (2000). A New Cloud Physics Parameterization in a Large-Eddy Simulation Model of Marine Stratocumulus. *Monthly Weather Review*, *128*(1). https://doi.org/10.1175/1520-0493(2000)128<0229:ANCPPI>2.0.CO;2

Kogan, Y. L., & Belochitski, A. (2012). Parameterization of Cloud Microphysics Based on Full Integral Moments. *Journal of the Atmospheric Sciences*, *69*(7), 2229–2242. https://doi.org/10.1175/JAS-D-11-0268.1

Lamb, K. D., Walqui, M. v.-L., Santos, S. P., & Morrison, H. (2023). *Reduced Order Modeling for Linearized Representations of Microphysical Process Rates* (preprint). Preprints. https://doi.org/10.22541/essoar.168995384.44033471/v1

Long, A. B. (1974). Solutions to the Droplet Collection Equation for Polynomial Kernels. *Journal of the Atmospheric Sciences*, *31*(4), 1040–1052. https://doi.org/10.1175/1520-0469(1974)031<1040:STTDCE>2.0.CO;2

Morrison, H., Kumjian, M. R., Martinkus, C. P., Prat, O. P., & van Lier-Walqui, M. (2019). A General N-Moment Normalization Method for Deriving Raindrop Size Distribution Scaling Relationships. *Journal of Applied Meteorology and Climatology*, *58*(2), 247–267. https://doi.org/10.1175/JAMC-D-18-0060.1

Morrison, H., van Lier-Walqui, M., Kumjian, M. R., & Prat, O. P. (2020). A Bayesian Approach for Statistical–Physical Bulk Parameterization of Rain Microphysics. Part I: Scheme Description. *Journal of the Atmospheric Sciences*, *77*(3), 1019–1041. https://doi.org/10.1175/JAS-D-19-0070.1

Morrison, H., van Lier-Walqui, M. v., Fridlind, A. M., Grabowski, W. W., Harrington, J. Y., Hoose, C., Korolev, A., Kumjian, M. R., Milbrandt, J. A., Pawlowska, H., Posselt, D. J., Prat, O. P., Reimel, K. J., Shima, S.-I., Diedenhoven, B. v., & Xue, L. (2020). Confronting the Challenge of Modeling Cloud and Precipitation Microphysics. *Journal of Advances in Modeling Earth Systems*, *12*(8). https://doi.org/10.1029/2019MS001689
e2019MS001689 2019MS001689.

Pinsky, M., Khain, A., & Shapiro, M. (2001). Collision Efficiency of Drops in a Wide Range of Reynolds Numbers: Effects of Pressure on Spectrum Evolution. *Journal of the Atmospheric Sciences*, *58*(7), 742–764. https://doi.org/10.1175/1520-0469(2001)058<0742:CEODIA>2.0.CO;2

Pruppacher, H., & Klett, J. (2010). Diffusion Growth and Evaporation of Water Drops and Snow Crystals | SpringerLink. In *Microphysics of Clouds and Precipitation* (pp. 502–567). Springer Netherlands. Retrieved May 28, 2024, from https://doi.org/10.1007/978-0-306-48100-0_13

Rodríguez Genó, C. F., & Alfonso, L. (2022). Parameterization of the collision–coalescence process using series of basis functions: COLNETv1.0.0 model development using a machine learning approach. *Geoscientific Model Development*, *15*(2), 493–507.

Rogers, R. (1975). An elementary parcel model with explicit condensation and supersaturation. *Atmosphere*, *13*(4), 192–204. https://doi.org/10.1080/00046973.1975.9648397

Seifert, A., & Beheng, K. D. (2006). A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 1: Model description. *Meteorology and Atmospheric Physics*, *92*(1-2), 45–66.

Seifert, A., & Rasp, S. (2020). Potential and Limitations of Machine Learning for Modeling Warm-Rain Cloud Microphysical Processes. *Journal of Advances in Modeling Earth Systems*, *12*(12), e2020MS002301. https://doi.org/https://doi.org/10.1029/2020MS002301

Shima, S.-I., Kusano, K., Kawano, A., Sugiyama, T., & Kawahara, S. (2009). The super-droplet method for the numerical simulation of clouds and precipitation: A particle-based and probabilistic microphysics model coupled with a non-hydrostatic model. *Quarterly Journal of the Royal Meteorological Society*, *135*(642), 1307–1320. https://doi.org/10.1002/qj.441

Shipway, B. J., & Hill, A. A. (2012). Diagnosis of systematic differences between multiple parametrizations of warm rain microphysics using a kinematic framework. *Quarterly Journal of the Royal Meteorological Society*, *138*(669), 2196–2211. https://doi.org/10.1002/qj.1913

Unterstrasser, S., Hoffmann, F., & Lerch, M. (2017). Collection/aggregation algorithms in Lagrangian cloud microphysical models: Rigorous evaluation in box model simulations. *Geoscientific Model Development*, *10*(4), 1521–1548. https://doi.org/10.5194/gmd-10-1521-2017

Chapter 5

WHAT CAN LIGHTNING AND SHIPPING REGULATIONS TELL US ABOUT AEROSOLS IN DEEPLY CONVECTING CLOUDS?

de Jong, E. K., Schneider, T., Jalkanen, J.-P., & Wennberg, P. O. (2024). What can lightning and shipping regulations tell us about aerosols in deeply convecting clouds? *ESS Open Archive; Submitted to Geophysical Research Letters*. https://doi.org/10.22541/essoar.171632547.79526166/v1

Abstract

Lightning flashes result from charge separation caused by colliding ice particles. They are particularly prevalent in continental convection but also occur over oceans, where their frequency correlates with maritime shipping emissions. In 2020, the International Maritime Organization (IMO) restricted ship fuels to much lower sulfur concentrations. We use this abrupt perturbation to study the response of oceanic lightning to aerosol, cloud, and convective properties. In principle, a reduction in aerosols should lead to a decrease in lightning over shiptracks relative to unpolluted regions. However, we find only one of the two investigated shipping lanes displays this sensitivity. Using the spatiotemporal sensitivity of lightning flash rate to hydrometeor size, we estimate the sensitivity of cloud water path to cloud number concentration. The results offers insights into both the role of aerosols in generating oceanic lightning and the mechanisms of aerosol-cloud interactions in deep convection.

Plain Language Summary

Human-emitted aerosols influence clouds in complex ways, which can be challenging to measure directly. We study the ways in which aerosols can impact the prevalence of lightning over the ocean by examining regions where shipping emissions have recently decreased. Using observational data about clouds, winds, and aerosols in this region, we demonstrate that aerosols generate lightning by producing more numerous small cloud ice particles, and we estimate the magnitude of aerosols' impact on these thunderstorm clouds.

5.1 Introduction

Aerosol-cloud interactions (ACI) remain the largest uncertainty among anthropogenic forcers of climate change (Intergovernmental Panel on Climate Change, 2021) due to complex feedbacks between cloud condensation nuclei (CCN) and cloud properties such as top-of-atmosphere reflectivity or cloud lifetime effects. For instance, the scientific community disagrees on whether increased CCN loading from aerosols in a cloud can invigorate convective systems (e.g. Fan et al., 2018; A. L. Igel and van den Heever, 2021; Rosenfeld et al., 2008; Varble et al., 2023). Furthermore, well-quantified ACI impacts such as cloud brightening (Twomey, 1977) and precipitation suppression (Yuan et al., 2023) have been predominantly studied in stratocumulus clouds, using satellite-based measurements over polluted shiptracks, but a broader characterization remains lacking.

In particular, our understanding of ACI in deep convective clouds (DCCs) remains incomplete, in part because satellite retrievals of aerosol and near-surface atmospheric properties in convective clouds are challenging (Koren et al., 2010). Our current understanding of ACI in this regime relies primarily on modeling studies, with contradictory results. For instance, Leung and van den Heever, 2023 and Dagan, 2024 find that aerosols can feed back onto atmospheric circulations, while Blossey et al., 2018 find a negligble net radiative effect from ACI over a polluted shipping lane in the Indian Ocean. The correlation between aerosol-optical depth (AOD) and markers of convection such as cloud top height or cloud morphology often confounds observational studies, rendering their results difficult to interpret (M. R. Igel & van den Heever, 2015; Koren et al., 2010; Zamora & Kahn, 2020).

Here we present an alternative method for quantifying and understanding aerosol impacts in convective clouds using a different marker of deep convection: lightning. Thornton et al., 2017 noted an increased frequency of oceanic lightning over two of the busiest shipping lanes in the Indian Ocean (IO) and the South China Sea (SCS). At the time, this finding was interpreted to support hypotheses of convective invigoration by aerosols (Chang, 2017). However, more recent works suggest that other mechanisms may be the source of this striking spatial correlation (Grabowski & Morrison, 2020; Peterson, 2023). An opportunity to revisit these results comes from the 2020 IMO mandate to reduce the maximum sulfur content in shipping fuels from 3.5% to 0.5% (IMO-2020). This resulted in reduced shipping-related emissions globally (Yuan et al., 2022) and detectable changes in low clouds (Diamond, 2023), with yet-unquantified impacts on convective clouds. The combination of spatial and temporal variation in emissions presents a unique opportunity to investigate the relationship between lightning, cloud microphysics, and aerosols in the IO and SCS deep convective regions. Using observational and reanlaysis data, we examine the spatio-temporal correlations between lightning, hydrometeors, meteorology, and convection, to provide new observational constraints on the magnitude of aerosol-cloud interactions in DCCs.

5.2 Theoretical basis

Lightning flashes stem from charge separation in clouds. This separation is induced by transfer of charge during ice collisions: large graupel particles grow through collisions, acquiring negative charge, and eventually fall, while lighter ice particles take on positive charge during collisions and are lofted (Takahashi et al., 1999). The mechanism relies on two key ingredients for lightning: ice particles and updrafts, which have strong meteorological co-variability. Theoretically, a higher concentration of heterogeneously-sized ice particles would lead to greater charge separation, as would a stronger updraft velocity. Both of these effects have been documented through correlations of lightning with convective indicators (Romps et al., 2018) and with the ice particle effective diameter (Sherwood et al., 2006).

Both cloud ice and convective strength are linked to aerosol loading. A large concentration of CCN or ice-nucleating particles (INP) increases the cloud particle concentration given a fixed cloud water content, leading to smaller effective radii (the Twomey effect) (Twomey, 1977). A stronger updraft velocity can similarly nucleate or freeze more ice crystals, which may lead to convective invigoration (Rosenfeld et al., 2008). The interplay of aerosols and convection is confounded by interactions with meteorological variables such as wind shear (M. R. Igel & van den Heever, 2015), leading to challenges evaluating the sign or magnitude of aerosol-cloud-lightning interactions. However, by considering several years of measurements including convective indicators such as sea surface temperature (SST), we isolate these meteorological factors from aerosol-induced changes in lightning and cloud microphysics.

5.3 Methods

5.3.1 Data

We investigate the same IO and SCS regions as Thornton et al., 2017, using lightning flash rate (LFR) data from the Earth Networks Total Lightning Network (Zhu et al., 2022). We focus on the period of 2017–2022, during which interannual variability in lightning is relatively low (Figure 5.1). In addition, we use observational and modeled variables, grouped into five categories and listed in Table 5.1. They include modeled shipping emission estimates from 2018–2021 produced using the Ship Traffic Emissions Assessment Model (STEAMv4.3.1) (Jalkanen et al., 2009), with weather effects disabled. As the LFR was previously shown to be spatially sensitive to shipping emissions during the cold season (Thornton et al., 2017), we restrict our analysis to November–April (IO) and October–March (SCS).

The LFR includes both cloud-to-ground and intracloud lightning. The choice to use vertical pressure velocity, SST, and horizontal wind shear reflects the finding of M. R. Igel and van den Heever, 2015 that these three quantities are stronger predictors of convective cloud properties than convective available potential energy (CAPE). Shear is computed as the difference between the 300–100hPa and 1000–900hPa wind vectors.

Data are interpolated to a 0.1° grid and investigated as monthly averages to reduce high-frequency variability. In the correlation study (section 4.2), we consider two regional subsets of this data: an IO region bounded by (1°N, 82°E) to (11°N, 93°E), and an SCS region bounded by (1.5°N, 104°E) and (11°N, 114°E), excluding continental gridboxes. Within each ocean basin, we further identify a "shipping lane" and "reference region" in Figures 5.1 and 5.2. The shipping lane domain is identified by the latitude/longitude ray of the highest density of shipping emissions, plus 1.5° perpendicular in each direction. The reference lane is defined by translating the shipping lane domain by 4° toward the open ocean away from heavy shipping activity.

5.3.2 Regression and correlation

For the LFR at a particular location y and time t, we assume a power-law dependence on the variables in Table 5.1:

$$L(y,t) = L_0 + \lambda \prod_i x_i(y,t)^{\gamma_i}. \tag{5.1}$$

Here, $L_0 > 0$ represents a constant baseline level of lightning, x_i are the normalized independent variables, and γ_i are the sensitivities. The normalization of a variable X_i is defined as

$$x_i(y,t) = \frac{X_i(y,t) - \min_{y,t}(X_i(y,t))}{\bar{X}_i - \min_{y,t}(X_i(y,t))}, \tag{5.2}$$

Category	Variable (abbreviation) [units]	Source
Lightning	Flash Rate [km−2 mo^{-1}]	Earth Networks Total Lightning Network (Zhu et al., 2022)
Aerosol loading	Shipping SO$_x$ emissions (SOx) [kg km−2 mo^{-1}]	STEAM v4.3.1 (Jalkanen et al., 2009; Johansson et al., 2017)
	2.5 μm Particulate Matter (pm25) [kg m^{-3}]	MERRA-2 (Global Modeling and Assimilation Office (GMAO), 2015a)
	Aerosol Optical Depth (AOD) [-]	MODIS (Science Investigator-led Processing System, n.d.)
	Fine Mode Fraction (FMF) [-]	MODIS (Science Investigator-led Processing System, n.d.)
Cloud properties	Mean ice particle diameter (D_ice_mean) [μm]	MODIS (Pincus & Hubanks, 2022)
	Mean liquid particle diameter (D_liq_mean) [μm]	MODIS (Pincus & Hubanks, 2022)
	Std dev ice particle diameter (D_ice_std) [μm]	MODIS (Pincus & Hubanks, 2022)
	Std dev ice particle diameter (D_liq_std) [μm]	MODIS (Pincus & Hubanks, 2022)
	Ice cloud water path (CWP_ice) [g m−2]	MODIS (Pincus & Hubanks, 2022)
	Liquid cloud water path (CWP_liq) [g m−2]	MODIS (Pincus & Hubanks, 2022)
	Cloud top pressure (Cloud_Top_Pressure) [mbar]	MODIS (Pincus & Hubanks, 2022)
Convection	Vertical pressure velocity (omega500) [Pa s^{-1}]	MERRA-2 (Global Modeling and Assimilation Office (GMAO), 2015c)
	Surface precipitation rate (precip) [mm hr^{-1}]	IMERG (Huffman et al., 2023)
	Cloud-top temperature (CTT) [K]	CERES (NASA/LARC/SD/ASDC, 2017)
	Cloud-top height (CTH) [km]	CERES (NASA/LARC/SD/ASDC, 2017)
Background and meteorology	950 hPa windspeeds (u950, v950) [m s^{-1}]	MERRA-2 (Global Modeling and Assimilation Office (GMAO), 2015c)
	Horizontal wind shear (shear) [m s^{-1}]	MERRA-2 (Global Modeling and Assimilation Office (GMAO), 2015c)
	Sea-surface skin temperature (SST) [K]	MERRA-2 (Global Modeling and Assimilation Office (GMAO), 2015b)

Table 5.1: Data used listed by category, including the relevant abbreviation and units corresponding to Figure 5.3. Some data resources are purely observational (lightning, MODIS, IMERG, CERES), while others are model or reanalysis products from MERRA-2 or STEAM.

where $\bar{X}_i$ is the mean value over space and time. We likewise define a normalized LFR $\mathcal{L}$,

$$\mathcal{L} = \frac{L(y,t) - L_0}{\lambda}, \tag{5.3}$$

and re-express Eq. 5.1 as a log-linear model

$$\log \mathcal{L} = \sum_i \gamma_i \log x_i, \tag{5.4}$$

with the sensitivities γ_i determined by linear regression. We additionally consider "feature importance", defined as the coefficients from ridge regression using all variables.

5.4 Results

5.4.1 Impacts of IMO-2020 on Lightning

Following IMO-2020, ship-emitted SO_x dropped by 80% over major shipping lanes passing through the port of Singapore, including the IO and SCS shipping regions studied by Thornton et al., 2017 (Figures 5.1 and 5.2). In comparison, changes to LFR over the same region are less obvious. Both shipping lanes display a reduced but still visually apparent local enhancement of LFR after 2020 (Figure 5.2, middle left). However, while the SCS shipping lane shows a distinct reduction in the normalized strike density, the IO shipping lane displays no apparent LFR trend following the regulation (Figure 5.2, bottom left). This difference is apparent in Figure 5.1, where the difference between the SCS shipping lane and reference LFR is reduced following 2020, whereas the IO shipping lane lightning displays stronger variability and no clear trend relative to the adjacent reference region. The Strait of Malacca, like the SCS shipping region, displays a dramatic reduction in both emissions and LFR across 2020, but with no adjacent shipping-free oceanic reference region for comparison.

5.4.2 Feature Importance and Selection

Figure 5.3 shows the regional linear feature importance of all normalized log-space variables in determining the spatiotemporal patterns of lightning. We investigate both monthly data and cold-season averages to illustrate differences between seasonal and interannual variations. For instance, this distinction illustrates the importance of seasonal variations in convective strength through stronger SST contributions in monthly versus annual data. The contrast further elucidates differences between controlling factors of lightning in the two basins. Microphysical variables including the cloud water path (CWP) are strong predictors of the annual lightning

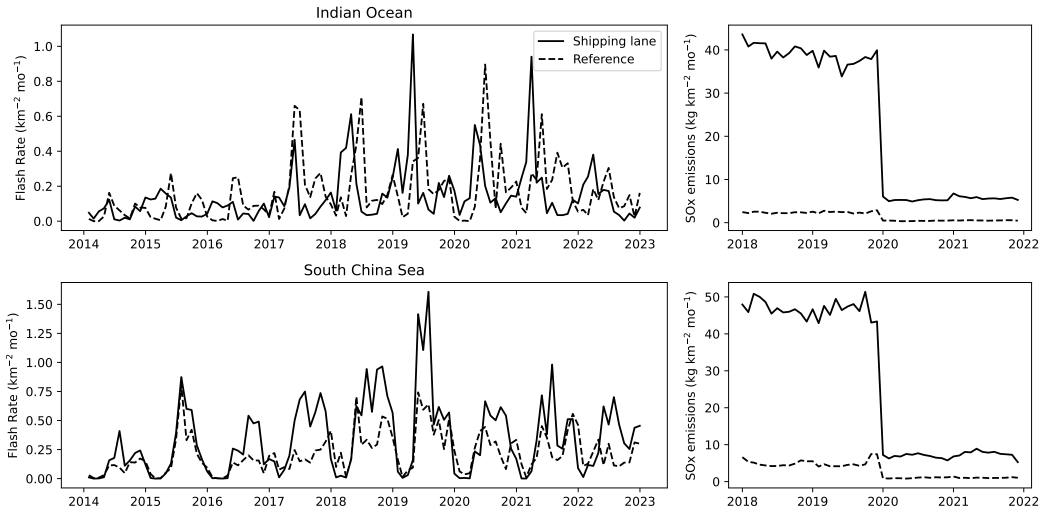


Figure 5.1: Timeseries of LFR (left) and SO_x emissions rate (right) over regions of the Indian Ocean and South China Sea. The solid line corresponds to the shipping lane region, while the dashed lines correspond to the adjacent reference regions.

intensity over IO, whereas these variables have a stronger predictive ability for the monthly lightning intensity over SCS.

Measurements of aerosol loading have similar feature importance in both ocean basins, with AOD and emissions being the best predictors in the relatively pristine IO basin, and with pm2.5 and Fine-Mode-Fraction (FMF) having a stronger correlation in the more polluted and near-continental SCS region. Indicators of convective strength do not have significantly larger feature importance than meteorological variables in the IO, indicating a microphysics-driven response of lightning to aerosol. However, these convective indicators play a stronger role than microphysical variables in the SCS, suggesting that convective drivers may outpace the influence of aerosols in this region.

Based on the feature evaluation described above, we downselect to 6 variables for the multivariate regression on annual aggregates described by Equation 5.4. These variables are the best representatives from the aerosol loading, cloud properties, and background/meteorology feature categories (Table 5.1), including: SO_x emissions and FMF, which correlate best with the IO and SCS regions, respectively; the mean ice and liquid particle diameters, which are the most directly computed MODIS satellite products; and shear and SST, which are selected by the automated regression techniques. No variables from the convection category are used, as predictors from this category are strongly correlated with the selected background variables and were

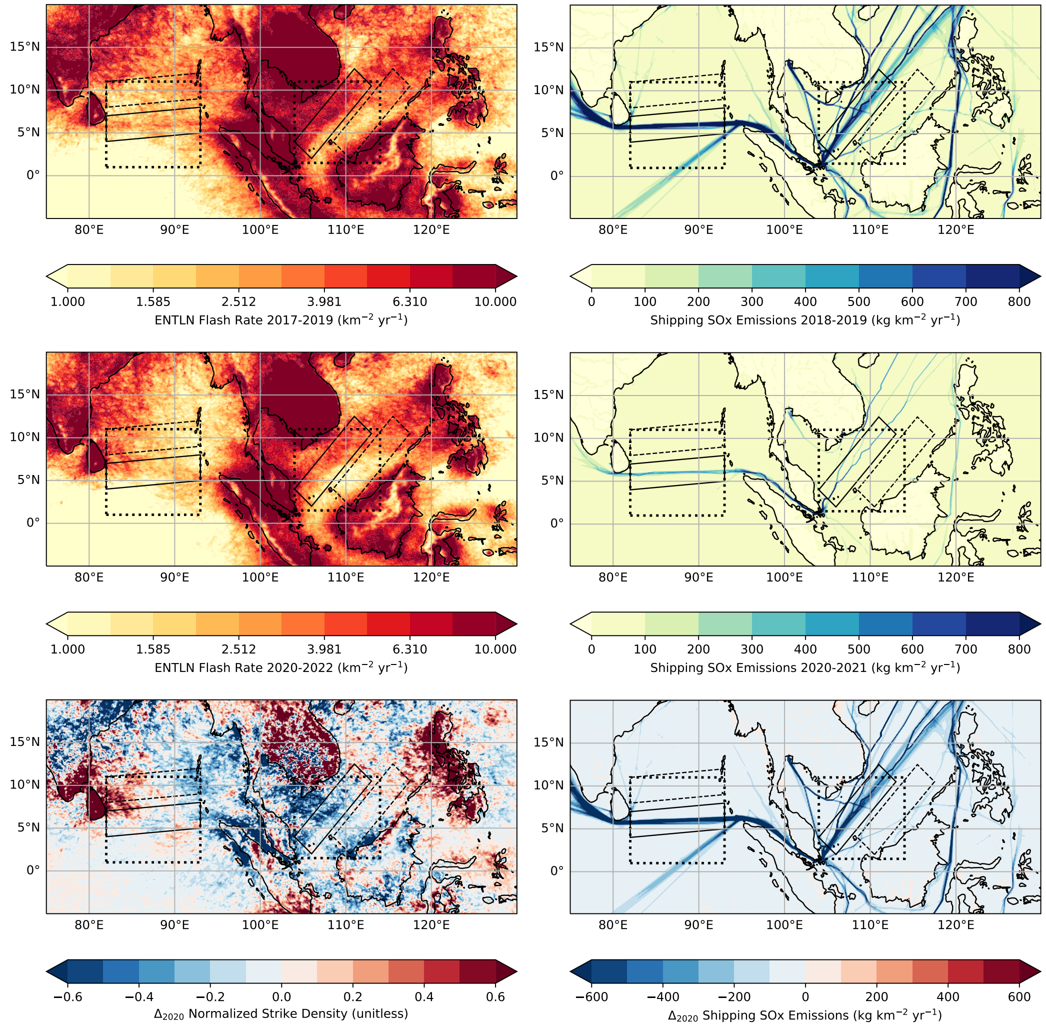


Figure 5.2: LFR (left column) and SO_x shipping emissions rate (right column) over the Indian Ocean and South China Sea prior to (top row) and following (middle row) IMO-2020 . Bottom row depicts the normalized difference in annual strike density (left) and the absolute difference in annual SO_x shipping emissions (right). In each panel, the solid box bounds the two shipping lanes of interest, the adjacent dashed boxes bound reference regions with less shipping activity, and the dotted boxes outline the larger regions analyzed in the correlation study.

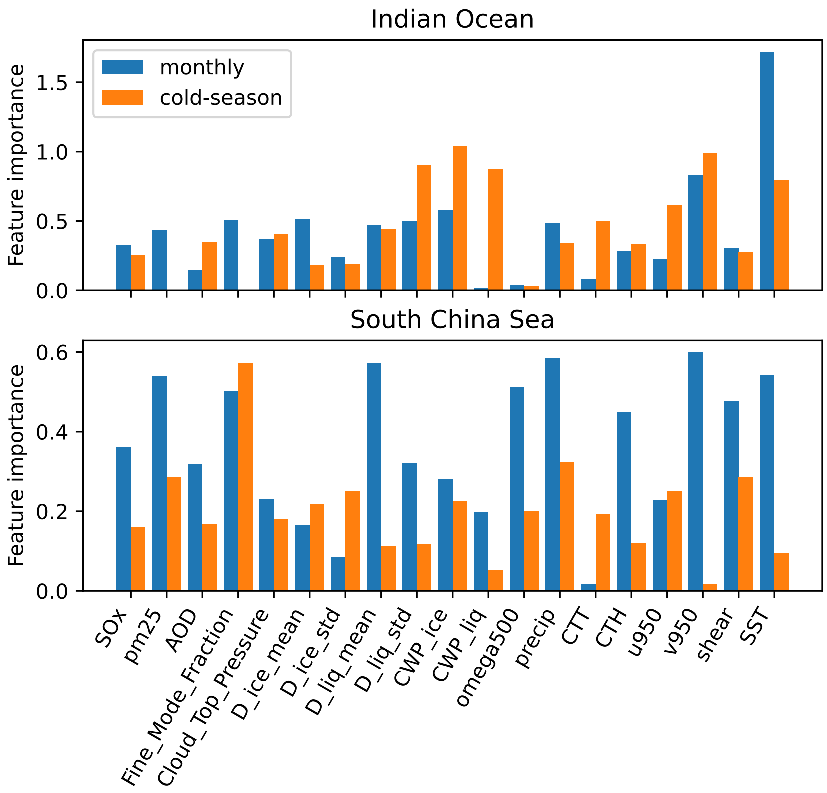


Figure 5.3: Feature importance for the normalized log-basis independent variables (Equation 5.2) over the IO region (top) and SCS (bottom).

rejected by both automated modeling algorithms. The coefficient of determination for this model structure is $R^2 = 0.56$ and $R^2 = 0.28$ for the SCS and IO regions, respectively (see Figure 5.4). Our regression model compares well to the trivariate regressions of Romps et al., 2018, which attained $R^2 < 0.3$.

5.4.3 Relationship of lightning to aerosol, microphysical, and background properties

The sensitivities of LFR to the independent variables are listed in Table 5.2 and depicted individually as functions of latitude (or distance from the primary shipping lane) and year in Figure 5.5. The spatial pattern of SCS lightning strikes is well-predicted following IMO-2020, while the LFR is generally underestimated pre-2020 in the region of strongest emissions. In the IO, our model generally underpredicts LFR and overpredicts the reduction in peak lightning at 6°N following IMO-2020. These predictive deficiencies suggest that lightning over the IO shipping lane is not aerosol-limited, and is therefore less sensitive to shipping emissions.

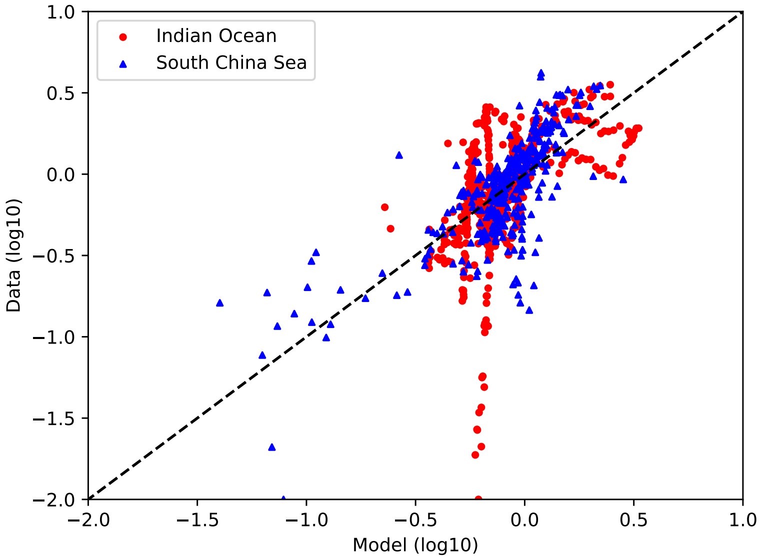


Figure 5.4: Model parity for the hand-selected feature model. $R^2 = 0.28$ for the Indian Ocean region and $R^2 = 0.56$ for the South China Sea region.

Variable x_i	IO Sensitivity γ_i	SCS Sensitivity γ_i
SO_x Emissions	0.27 ± 0.06	0.34 ± 0.07
Fine Mode Fraction	0.13 ± 0.05	0.71 ± 0.04
Mean ice particle size	0.06 ± 0.04	-0.35 ± 0.06
Mean liquid particle size	-0.22 ± 0.06	0.56 ± 0.08
Shear	-0.27 ± 0.03	-0.27 ± 0.08
SST	0.44 ± 0.07	0.11 ± 0.04

Table 5.2: Coefficients of the normalized log-space linear regression, shown as value plus/minus one standard error.

5.4.3.1 Aerosol loading

Both shipping regions have a positive sensitivity to sulfur emissions and FMF, with a particularly strong sensitivity of SCS lightning to FMF (Table 5.2). Strong sensitivity of lightning to ultrafine aerosols ($< 1\mu m$) supports previous findings of a contrasting relationship of lightning to fine versus coarse sea salt aerosol (Pan et al., 2022). Sensitivity to the location rather than magnitude of emissions supports the hypothesis that ships themselves attract lightning (Peterson, 2023). This hypothesis appears more likely in the IO case, where the lightning peak is relatively consistent before and after emission regulations. The FMF peak over each shipping lane indicates enhanced fine aerosol concentrations, and the mean reduction in fine aerosols following IMO-2020 indicates contributions from shipping emissions (Figure 5.5, 2nd row). Additional factors such as the local winds may distribute and/or concentrate aerosols. For instance, the IO meridional near-surface wind (v950, not shown) has a local minimum at 6°N and a sign-change around this latitude in the cold-season mean, and therefore may concentrate aerosols near the IO shipping lane regardless of their origin.

5.4.3.2 Microphysical properties

In the years following IMO-2020, the average ice particle size decreased over both shipping regions by up to $1 \mu m$, while the average liquid particle size increased (Figure 5.5, rows 3-4). According to the model sensitivities (Table 5.2), lightning has a contrasting response to hydrometeor size in the two shipping regions. IO lighting has a negligible sensitivity to ice particle size and a negative relationship to liquid particle size. The liquid particle diameter additionally shows a small local minimum at 6°N, where aerosol emissions are concentrated, particularly pre-2020, which indicates a warm-phase Twomey effect. SCS lightning is more strongly correlated with both phases of hydrometeors, with a negative relationship to ice particle diameter (Liu et al., 2020; Sherwood et al., 2006) and a positive relationship to liquid particle diameter. The negatively correlated SCS ice particle size shows a substantial dip near the location of shipping emissions, indicating a Twomey-like effect in the cold-phase. The strong positive relationship of liquid particle diameter and SCS lightning is more puzzling. One potential explanation is a cycle of aerosol-induced precipitation suppression contributing to cold-phase convective invigoration, followed by more intense precipitation of large liquid hydrometeors alongside stronger electrification. However, the liquid hydrometeor size shows little interannual variability relative to other microphysical quantities and may have been unimpacted by changes to aerosol emissions following IMO-2020.

5.4.3.3 Background and meteorology

Lightning has a nearly identical sensitivity to wind shear in both oceanic shipping regions. The wind shear in turn displays the largest interannual variability among the model variables, demonstrating the strong role of meteorology and horizontal winds in driving convection over the ocean. The positive correlation between LFR and SST follows the expected relationship of SST and convection (M. R. Igel & van den Heever, 2015) and contributes to the interannual variability.

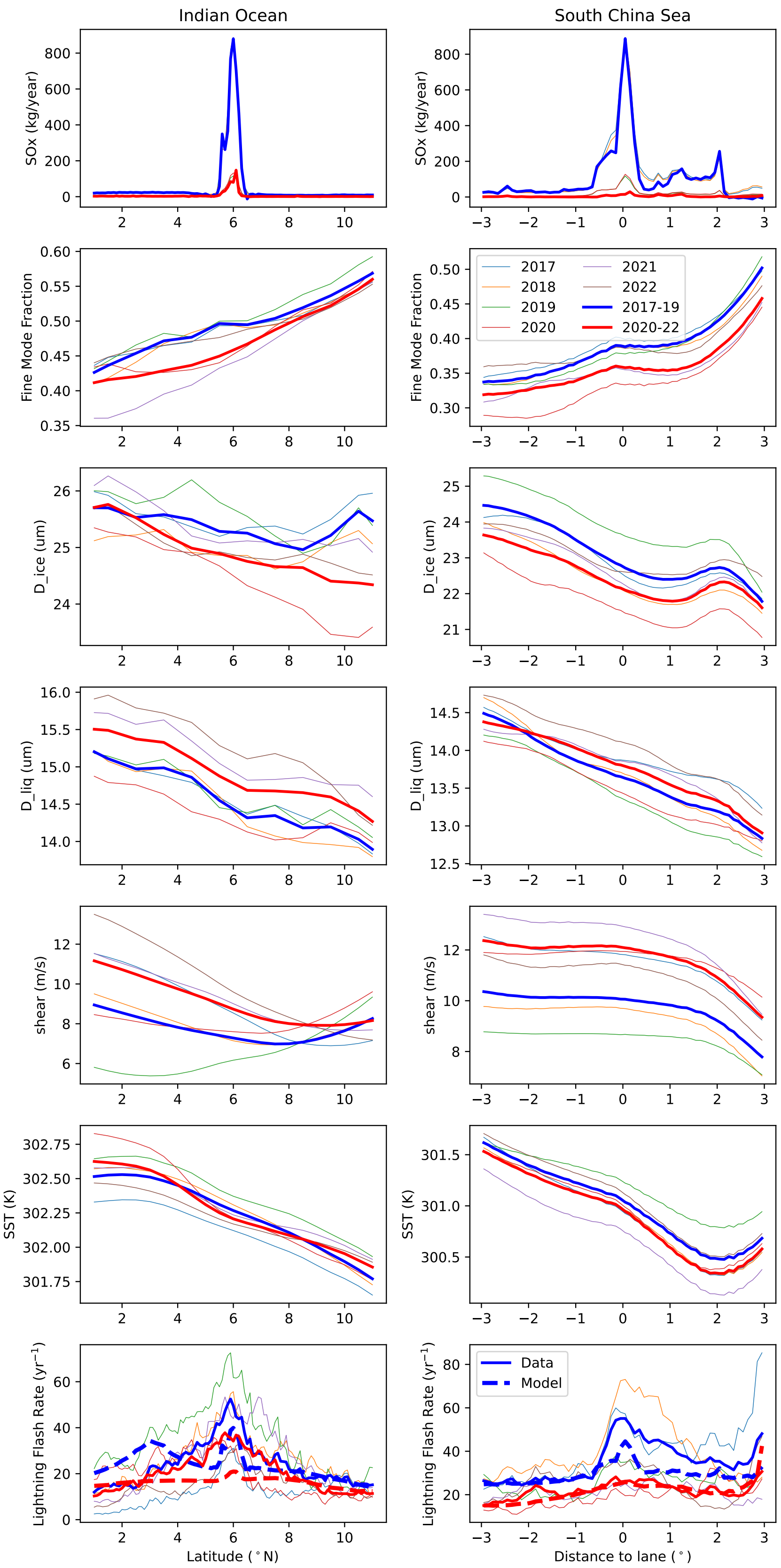


Figure 5.5: Cold-season mean model features as a function of latitude (or distance from shipping lane) and year (top 6 rows) and the overall model predictions versus LFR (bottom row). Pre- and post-IMO-2020 means are depicted in thicker blue and red lines, respectively.

5.5 Interpretation of microphysical sensitivities

Using a simple mechanistic explanation of lightning, we now attempt to isolate and compare the relative strength of Twomey-like and non-Twomey contributions of aerosol loading to lightning and cloud properties. We begin with the limiting assumption that LFR is proportional to the ice hydrometeor collision rate C, which in turn scales as

$$\mathcal{L} \propto C \propto N^2 r^\alpha, \tag{5.5}$$

where r and N are the effective radius and number concentration of hydrometeors, respectively.

Assuming a fixed CWP and cloud fraction, $N \propto r^{-3}$, leads to the result that $\mathcal{L} \propto C \propto r^{\alpha-6}$ for a strictly Twomey-type aerosol-cloud-lightning interaction. A purely gravitational collision has a coefficient of $\alpha = 4$, while a turbulent collisional kernel such as the Saffman-Turner kernel (Saffman & Turner, 1956) uses $\alpha = 3$. The normalized sensitivity of lightning to ice particle size in both ocean basins ($\gamma_i \in (-0.35, 0.06)$) implies that $\alpha - 6 = -0.15 \pm 0.2$, i.e., $\alpha \approx 6$, which falls outside the noted range for gravitational or turbulent collisions. Thus, it is unlikely that this dependence is purely due to the Twomey effect. Under a contrasting assumption of fixed cloud particle concentration (and variable CWP), $\mathcal{L} \propto r^\alpha$, with the data implying $\alpha \approx 0$. The aforementioned realistic value of $\alpha = 3.5 \pm 0.5$ imply a collision rate dependence of $C \propto N^2 r^\alpha = r^{\alpha-\beta}$, i.e., $N \propto r^{-\beta/2}$ with $\beta = 3.6 \pm 0.7$. Retaining the assumption of fixed cloud fraction but now allowing both number concentration and CWP to vary, we obtain that

$$\mathrm{CWP} \propto N r^3 \propto N^{1-6/\beta}. \tag{5.6}$$

This relationship indicates that

$$\frac{\partial \log \mathrm{CWP}}{\partial \log N} = 1 - 6/\beta \approx -0.7 \pm 0.3, \tag{5.7}$$

providing a novel estimate for the sensitivity of cloud water path to hydrometeor concentration (and by proxy, the concentration of CCN and INP). This derivation extends the language and logic of Bellouin et al., 2020 and Yuan et al., 2023 to constrain aerosol-cloud interactions in DCCs by using lightning as an indirect measurement of the combined effects of number concentration and CWP.

This preliminary estimate, though uncertain, supports a model of pristine deep clouds that are highly sensitive to aerosol concentration. Without the competing low-cloud mechanism of entrainment drying (Yuan et al., 2022), we find that these DCCs are susceptible to precipitation suppression and potentially convective invigoration. Aerosols introduced by ship traffic therefore intensify electrification by generating wetter, deeper, and/or more persistent clouds.

5.6 Discussion

Using the LFR signal in response to a sharp perturbation of shipping emissions, this study quantitatively assessed the relationship between lightning, aerosol emissions, cloud microphysical properties, and background meteorology. Unlike previous studies that correlated continental lightning with convective metrics, we find that oceanic lightning is best explained as a response of cloud microphysics to local aerosol loading on top of a background convective state that is influenced by meteorological and surface conditions. This combination of rapid and slow responses is supported by the difference in feature importance between monthly and seasonal data aggregates in explaining LFR.

While our linear model cannot precisely recover the spatiotemporal distribution of lightning over the Indian Ocean shipping lane, it has strong predictive power over the SCS and reveals the mechanisms responsible for a visually apparent increase in lightning over both busy shipping regions. We find a local peak in aerosol fine-mode-fraction that corresponds to the location of heaviest shipping emissions, which in turn influences the spatial distribution of lightning in both regions. We additionally find that the SCS shipping region follows an expected negative relationship between LFR and ice particle size, supporting a cold-phase Twomey-like effect that stems from locally enhanced aerosol concentration. The relationships of lightning and hydrometeor size over the Indian Ocean appear much weaker, suggesting that the IO lightning signal is not aerosol-limited but susceptible to other factors such as meteorology.

Finally, we suggest a simple estimate for the sensitivity of cloud water path to number concentration in this deeply convecting region by using the relationship of LFR to hydrometeor size. Similar constraints on the sensitivity of CWP to number concentration are restricted to liquid-phase low clouds Yuan et al., 2023, e.g. and cannot be expected to encompass complex feedbacks such as convective invigoration (or enervation) at work in mixed phase clouds. Our estimate of CWP sensitivity is inherently uncertain due to assumptions of fixed cloud fraction and lightning as proportional to ice particle collision rate, yet it suggests a sensitivity that is more strongly positive than found by related low-cloud studies. Flawed as it may be, this procedure offers a step toward observationally constraining aerosol-cloud interactions that cannot be directly measured from space.

Open Research

Datasets used in this study are summarized in Table 5.1. MERRA-2 reanalysis products (Global Modeling and Assimilation Office (GMAO), 2015a, 2015b, 2015c) and satellite products (Huffman et al., 2023; NASA/LARC/SD/ASDC, 2017; Pincus & Hubanks, 2022; Science Investigator-led Processing System, n.d.) are publicly available through NASA databases at corresponding reference URL's. Lightning flash rate measurements are provided by Earth Networks Inc., an AEM company (Zhu et al., 2022). Shipping emissions inventory data was provided by the Finnish Meteorological Institute (Jalkanen et al., 2009). All analysis in this study employed open-source software tools in the Python programming language.

Acknowledgements

We are grateful for productive scientific discussions with Chris Wright, Ryan Ward, Dr. Michael Peterson, and Dr. Adele Igel. E. de Jong was supported by a Department of Energy Computational Sciences Graduate Fellowship. This research was additionally supported by Eric and Wendy Schmidt (by recommendation of Schmidt Futures). POW acknowledges support from the Simons Foundation.

References

Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., Stier, P., Watson-Parris, D., Boucher, O., Carslaw, K. S., Christensen, M., Daniau, A.-L., Dufresne, J.-L., Feingold, G., Fiedler, S., Forster, P., Gettelman, A., Haywood, J. M., Lohmann, U., Malavelle, F., Mauritsen, T., . . . Stevens, B. (2020). Bounding Global Aerosol Radiative Forcing of Climate Change. *Reviews of Geophysics*, *58*(1), e2019RG000660. https://doi.org/10.1029/2019RG000660

Blossey, P. N., Bretherton, C. S., Thornton, J. A., & Virts, K. S. (2018). Locally Enhanced Aerosols Over a Shipping Lane Produce Convective Invigoration but Weak Overall Indirect Effects in Cloud-Resolving Simulations. *Geophysical Research Letters*, *45*(17), 9305–9313. https://doi.org/10.1029/2018GL078682

Chang, S. (2017). Ships cause their own stormy seas. *Physics Today*, *70*(11), 20–21. https://doi.org/10.1063/PT.3.3755

Dagan, G. (2024). Large-Scale Tropical Circulation Intensification by Aerosol Effects on Clouds. *Geophysical Research Letters*, *51*(7), e2024GL109015. https://doi.org/10.1029/2024GL109015

Diamond, M. S. (2023). Detection of large-scale cloud microphysical changes within a major shipping corridor after implementation of the International Maritime Organization 2020 fuel sulfur regulations. *Atmospheric Chemistry and Physics*, *23*(14), 8259–8269. https://doi.org/10.5194/acp-23-8259-2023

Fan, J., Rosenfeld, D., Zhang, Y., Giangrande, S. E., Li, Z., Machado, L. A. T., Martin, S. T., Yang, Y., Wang, J., Artaxo, P., Barbosa, H. M. J., Braga, R. C., Comstock, J. M., Feng, Z., Gao, W., Gomes, H. B., Mei, F., Pöhlker, C., Pöhlker, M. L., . . . de Souza, R. A. F. (2018). Substantial convection and precipitation enhancements by ultrafine aerosol particles. *Science*, *359*(6374), 411–418. https://doi.org/10.1126/science.aan8461

Global Modeling and Assimilation Office (GMAO). (2015a). MERRA-2 instM_2d_asm_nx: 2d,Monthly mean,Single-Level,Assimilation,Single-Level Diagnostics V5.12.4, https://doi.org/10.5067/5ESKGQTZG7FO

Global Modeling and Assimilation Office (GMAO). (2015b). MERRA-2 instM_3d_asm_np: 3d,Monthly mean,Instantaneous,Pressure-Level,Assimilation,Assimilated Meteorological Fields V5.12.4. https://doi.org/10.5067/2E096JV59PK7

Global Modeling and Assimilation Office (GMAO). (2015c). MERRA-2 tavgM_2d_slv_nx: 2d,Monthly mean,Time-Averaged,Single-Level,Assimilation,Single-Level Diagnostics V5.12.4, https://doi.org/10.5067/AP1B0BA5PD2K

Grabowski, W. W., & Morrison, H. (2020). Do Ultrafine Cloud Condensation Nuclei Invigorate Deep Convection? *Journal of the Atmospheric Sciences*, *77*(7), 2567–2583. https://doi.org/10.1175/JAS-D-20-0012.1

Huffman, G., Stocker, E., Bolvin, D.T., Nelkin, E.J., & Jackson Tan. (2023). GPM IMERG Final Precipitation L3 1 month 0.1 degree x 0.1 degree V07. https://doi.org/10.5067/GPM/IMERG/3B-MONTH/07

Igel, A. L., & van den Heever, S. C. (2021). Invigoration or Enervation of Convective Clouds by Aerosols? *Geophysical Research Letters*, *48*(16), e2021GL093804. https://doi.org/10.1029/2021GL093804

Igel, M. R., & van den Heever, S. C. (2015). The relative influence of environmental characteristics on tropical deep convective morphology as observed by CloudSat. *Journal of Geophysical Research: Atmospheres*, *120*(9), 4304–4322. https://doi.org/10.1002/2014JD022690

Intergovernmental Panel on Climate Change. (2021). *Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change* (V. Masson-Delmotte, P. Zhai, A. Pirani, S. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou, Eds.). Cambridge University Press.

Jalkanen, J.-P., Brink, A., Kalli, J., Pettersson, H., Kukkonen, J., & Stipa, T. (2009). A modelling system for the exhaust emissions of marine traffic and its application in the baltic sea area. *Atmospheric Chemistry and Physics*, *9*(1), 9209–9223. https://doi.org/10.5194/acp-9-9209-2009

Johansson, L., Jalkanen, J.-P., & Kukkonen, J. (2017). Global assessment of shipping emissions in 2015 on a high spatial and temporal resolution. *Atmospheric Environment*, *167*, 403–415. https://doi.org/10.1016/j.atmosenv.2017.08.042

Koren, I., Feingold, G., & Remer, L. A. (2010). The invigoration of deep convective clouds over the Atlantic: Aerosol effect, meteorology or retrieval artifact? *Atmospheric Chemistry and Physics*, *10*(18), 8855–8872. https://doi.org/10.5194/acp-10-8855-2010

Leung, G. R., & van den Heever, S. C. (2023). Aerosol breezes drive cloud and precipitation increases. *Nature Communications*, *14*(1), 2508. https://doi.org/10.1038/s41467-023-37722-3

Liu, Y., Guha, A., Said, R., Williams, E., Lapierre, J., Stock, M., & Heckman, S. (2020). Aerosol Effects on Lightning Characteristics: A Comparison of Polluted and Clean Regimes. *Geophysical Research Letters*, *47*(9), e2019GL086825. https://doi.org/10.1029/2019GL086825

NASA/LARC/SD/ASDC. (2017). CERES-MODIS and hourly geostationary cloud properties stratified by ISCCP cloud types for day and night. https://doi.org/10.5067/TERRA+AQUA/CERES/CLDTYPHIST_L3.004

Pan, Z., Mao, F., Rosenfeld, D., Zhu, Y., Zang, L., Lu, X., Thornton, J. A., Holzworth, R. H., Yin, J., Efraim, A., & Gong, W. (2022). Coarse sea spray inhibits lightning. *Nature Communications*, *13*(1), 4289. https://doi.org/10.1038/s41467-022-31714-5

Peterson, M. J. (2023). *Interactions between Lightning and Ship Traffic* (preprint). Preprints. https://doi.org/10.22541/essoar.167870354.44683383/v1

Pincus, R., & Hubanks, P. (2022). MODIS Standard L3 MCD06 COSP Product. https://doi.org/10.5067/MODIS/MCD06COSP_M3_MODIS.061

Romps, D. M., Charn, A. B., Holzworth, R. H., Lawrence, W. E., Molinari, J., & Vollaro, D. (2018). CAPE Times P Explains Lightning Over Land But Not the Land-Ocean Contrast. *Geophysical Research Letters*, *45*(22), 12, 623–12, 630. https://doi.org/10.1029/2018GL080267

Rosenfeld, D., Lohmann, U., Raga, G. B., O'Dowd, C. D., Kulmala, M., Fuzzi, S., Reissell, A., & Andreae, M. O. (2008). Flood or Drought: How Do Aerosols Affect Precipitation? *Science*, *321*(5894), 1309–1313. https://doi.org/10.1126/science.1160606

Saffman, P. G., & Turner, J. S. (1956). "on the collision of drops in turbulent clouds". *Journal of Fluid Mechanics*, *1*(1), 16–30. https://doi.org/10.1017/S0022112056000020

Science Investigator-led Processing System. (n.d.). Visible Infrared Imaging Radiometer Suite NASA standard Level-3 monthly Deep Blue aerosol roducts. https://doi.org/10.5067/VIIRS/AERDB_M3_VIIRS_SNPP.002

Sherwood, S. C., Phillips, V. T. J., & Wettlaufer, J. S. (2006). Small ice crystals and the climatology of lightning. *Geophysical Research Letters*, *33*(5). https://doi.org/10.1029/2005GL025242

Takahashi, T., Tajiri, T., & Sonoi, Y. (1999). Charges on graupel and snow crystals and the electrical structure of winter thunderstorms. *Journal of the Atmospheric Sciences*, *56*(11). https://doi.org/10.1175/1520-0469(1999)056<1561:COGASC>2.0.CO;2

Thornton, J. A., Virts, K. S., Holzworth, R. H., & Mitchell, T. P. (2017). Lightning enhancement over major oceanic shipping lanes. *Geophysical Research Letters*, *44*(17), 9102–9111. https://doi.org/10.1002/2017GL074982

Twomey, S. (1977). The Influence of Pollution on the Shortwave Albedo of Clouds. *Journal of the Atmospheric Sciences*, *34*(7), 1149–1152. https://doi.org/10.1175/1520-0469(1977)034<1149:TIOPOT>2.0.CO;2

Varble, A. C., Igel, A. L., Morrison, H., Grabowski, W. W., & Lebo, Z. J. (2023). Opinion: A Critical Evaluation of the Evidence for Aerosol Invigoration of Deep Convection. *EGUsphere*, 1–31. https://doi.org/10.5194/egusphere-2023-938

Yuan, T., Song, H., Wood, R., Oreopoulos, L., Platnick, S., Wang, C., Yu, H., Meyer, K., & Wilcox, E. (2023). Observational evidence of strong forcing from aerosol effect on low cloud coverage. *Science Advances*, *9*(45), eadh7716. https://doi.org/10.1126/sciadv.adh7716

Yuan, T., Song, H., Wood, R., Wang, C., Oreopoulos, L., Platnick, S. E., von Hippel, S., Meyer, K., Light, S., & Wilcox, E. (2022). Global reduction in ship-tracks from sulfur regulations for shipping fuel. *Science Advances*, *8*(29), eabn7988. https://doi.org/10.1126/sciadv.abn7988

Zamora, L. M., & Kahn, R. A. (2020). Saharan Dust Aerosols Change Deep Convective Cloud Prevalence, Possibly by Inhibiting Marine New Particle Formation. *Journal of Climate*, *33*(21), 9467–9480. https://doi.org/10.1175/JCLI-D-20-0083.1

Zhu, Y., Stock, M., Lapierre, J., & DiGangi, E. (2022). Upgrades of the earth networks total lightning network in 2021. *Remote Sensing*, *14*. https://doi.org/10.3390/rs14092209

Chapter 6

CONCLUDING REMARKS

Chapters 2–5 of this dissertation are only one step of journey toward improving our understanding and models of cloud microphysics. In this concluding chapter, I wish to contextualize the challenges and methods addressed by earlier chapters within Atmospheric Science and broader disciplines, while offering suggestions and warnings for future work based on my experience.

6.1 State of the Science

Although collisional processes in warm rain microphysics present a substantial challenge to progress in climate and weather modeling, they are by no means the only such obstacle. Indeed, this thesis only utilizes a small subset of available modeling techniques and data resources to address an even smaller subset of underresolved atmospheric processes. Moreover, fields beyond atmospheric science have faced parallel challenges in parameterizing small scale processes. In fact, several of these fields have developed parallel modeling techniques whose adaptation to atmospheric processes could benefit the science.

Two important complexities of modeling cloud physics which are not fully discussed in this thesis include:

1. **Lack of Ground Truth:** In the case of most microscopic atmospheric processes including cloud microphysics, our ability to verify and validate models and parameterizations is hindered by a lack of directly comparable observations. Satellite observations often have too low of spatial and temporal resolution to sample microphysical processes, while aircraft and laboratory experiments are not necessarily a representative sample of diverse atmospheric states. Many processes such as secondary ice production are challenging to even reproduce in the laboratory, hindering our fundamental understanding. Yet other datasets, such as ground-based measurements of lightning, are proprietary or inaccessible for academic research.

2. **Parameterization-Squared:** Once a cloud microphysics model structure and its parameters are set, it must still be coupled to the representation of other

> subgrid scale processes. For instance, turbulent motions of the air are not fully resolved by weather and climate models, yet they impact cloud microphysics through fluctuations in the local water saturation and entrainment of dry air (Chandrakar et al., 2021). The tracers and parameters used to represent cloud microphysics must be further integrated with parameterizations of subgrid scale turbulence, often requiring yet further assumptions regarding how the parameterized quantities interact. When comparing model results with observations, it can be challenging to disentangle errors due to one subgrid-scale model from errors due to another, though the "piggybacking" approach (Grabowski, 2019) may aid in doing so.

Recent efforts to overcome these obstacles have focused on either improving model structure and parameters, with emphasis on integrate machine learning techniques, or on increasing model resolution. Techniques such as inverse problem solving (e.g. Bieli et al., 2022) offer a means of estimating the value and uncertainty of many parameters simultaneously. Taking ML one step further, other researchers attempt to replace ill-posed parameterizations with machine-learned surrogate models (e.g. Gettelman et al., 2021) in order to alleviate structural uncertainties. On the other hand, increasing the spatial resolution of climate models can eliminate the need for parameterizations of convective cloud motions, reducing the quantity of uncertain and entangled parameters at the cost of higher computational intensity. Neither ML nor model resolution solves the issue of the lack of a ground truth for microscopic processes, but each attempts to circumvent the need for such data by eliminating uncertain parameters and by bringing the scope and spatial scale of models closer to that of observations.

The models used in cloud microphysics are not unique to the subfield, but are often developed in isolation from other disciplines and in some cases face obstacles that have already been overcome by other scientific communities. For instance, Monte Carlo simulations of representative tracers (the Superdroplet method, in microphysics) has been in use for over a decade by astrophysicists (Zsom & Dullemond, 2008) and marine scientists (Jokulsdottir & Archer, 2016), and for nearly thirty years in the field of population dynamics (Scheffer et al., 1995). Bulk methods for representing a dispersed phase within fluid flow likewise exist in Computational Fluid Dynamics and combustion modeling to represent soot aggregates or other reactive particles. In particular, the Quadrature Method of Moments (McGraw, 1997) and more recent variants (e.g. Marchisio & Fox, 2005; Mueller et al., 2009) are an accurate and efficient class of bulk methods used for tracking sizes and even additional properties of particles, and have been successfully applied to aerosol formation and growth (Fierce & McGraw, 2017). These methods may outperform the alternative bulk and spectral representations presented in Chapters 3 and 4 and are worth pursuing further within cloud physics. Unfortunately, adapting and integrating new model components into large operational weather and climate models can be a laborious and ill-rewarded task. Thus even when new models are presented and validated in simplified contexts, such as in Chapters 2–4, this initial model development does not guarantee widespread adoption.

6.2 Next Steps and Words of Caution

6.2.1 High-Fidelity Modeling

Even with the inclusion of collisional breakup (Chapter 2), particle-based microphysics remains incomplete in its purview of microphysical processes. In particular, mixed-phase representations within the SDM (Shima et al., 2020) currently neglect secondary ice processes such as rime-splintering or collisional breakup of ice, as these processes lack substantial fundamental understanding. However, coupling the newly-introduced collisional breakup algorithm with SDM representations of immersion freezing (Arabas et al., 2024) could provide an effective way to represent the spontaneous freezing of supercooled liquid fragments upon collision. Exploring this SIP mechanism as a combination of two existing process rates could introduce a framework for representing and parameterizing additional collisional ice processes.

Although Chapters 3 and 4 utilize the SDM as a high-fidelity reference tool for bulk parameterizations and other studies utilize SDM simulations as training data for calibration (Azimi et al., 2023; Bieli et al., 2022), these practices are not yet widely-accepted in cloud physics. Indeed, the particle-based microphysics community is currently working to address discrepancies in precipitation predictions that arise from small variations in the representation of stochastic coalescence among realizations of the algorithm (Hill et al., 2023; Morrison et al., 2024). Other model parameters such as the model time step, horizontal resolution, and initialisation of superdroplets may additionally impact the results of an SDM simulation.

6.2.2 Simplified Models

The sequence of Chapters 3 and 4 traces the evolution of the flexible moment-based method, Cloudy.jl or "Cloudy", as a continuation and improvement upon the collocation method. Inspired by superdroplet model structure, Cloudy.jl was an attempt to maintain the flexibility and multimodal capability of a spectral method while resolving the issues of mass conservation that arose from the collocation method. Nevertheless, Cloudy is still not necessarily the best solution for representing warm rain microphysics, in particular due to its reliance on polynomial kernel approximations (which may be ill-behaved for large particles), numerical integration, and gamma or exponential type closures. For this reason, I suggest continuing to explore other model structures such as variants of the Quadrature Method of Moments (McGraw, 1997) alongside continued development of this flexible moment architecture. In particular, evaluation of Cloudy in more complex atmospheric simulations, such as LES reproducing results of the RICO field campaign, should be undertaken before coupling with mixed phase processes or a climate model. Future development of any bulk microphysics model, whether it be Cloudy, QMOM, or an entirely different structure, would benefit from the consideration of additional properties such as ice density or aspect ratio or aerosol hygroscopicity in order to unify the representation of microscopic particles.

The future of microphysics may not lie within this bucket of structured numerical models at all, but rather among data-driven models. The BOSS approach (Morrison et al., 2019) probed this possibility by abandoning structured process rates in favor of generalized power series and calibrated parameters, and is under development for large-scale simulations at the time of writing. Taken one step further, Lamb et al., 2023 found that the information of particle size distributions could be reduced to three inherent dimensions or latent variables using neural network encoders. Can ML techniques uncover not only this latent space, but also the governing equation for a complex system such as cloud microphysics? Using the tools developed by Lamb et al., 2023 and Champion et al., 2019, I intend to explore this question during my postdoctoral research using high-fidelity SDM simulations as a stand-in for observational data.

6.2.3 Observational Evidence

Compiling, downselecting, and processing observational data is a science of its own. Chapter 5 utilized only a small fraction of available datasets to study aerosol and lightning, yet even these data took nearly a year to compile, from requesting proprietary lightning measurements to running the STEAM model as an alternative to the EDGAR database used by Thornton et al., 2017. Downloading open-source and post-processed level-3 monthly satellite observations from NASA databases was easy by comparison. In fact, satellite products are available at much higher temporal frequency (as level-2 or level-1 datasets, rather than level-3) and could be used in a more detailed study to isolate the contributions of meteorology on daily lightning frequency. Inferencing techniques such as ship-track detection algorithms (Manshausen et al., 2022) and the comparison of counterfactuals using "kriging" (Diamond, 2023) could further strengthen the statistical evidence used to study aerosol-cloud-lightning interactions. Alternatively, bottom-up studies using lightning predictions from operational weather models could provide better mechanistic evidence for lightning production, even if the predictions themselves are imperfect. Ultimately, enhancing and upscaling observational data, whether through ML techniques or otherwise, may be the future of how we integrate observational evidence with models.

References

Arabas, S., Curtis, J. H., Silber, I., Fridlind, A. M., Knopf, D. A., West, M., & Riemer, N. (2024). Immersion freezing in particle-based aerosol-cloud microphysics: A probabilistic perspective on singular and time-dependent models. *arXiv preprint*. https://doi.org/10.48550/arXiv.2308.05015

Azimi, S., Jaruga, A., de Jong, E. K., Arabas, S., & Schneider, T. (2023). Training warm-rain bulk microphysics schemes using super-droplet simulations. https://doi.org/10.22541/essoar.169603562.20867617/v1

Bieli, M., Dunbar, O. R. A., de Jong, E. K., Jaruga, A., Schneider, T., & Bischoff, T. (2022). An Efficient Bayesian Approach to Learning Droplet Collision Kernels: Proof of Concept Using “Cloudy,” a New n-Moment Bulk Microphysics Scheme. *Journal of Advances in Modeling Earth Systems*, *14*(8), e2022MS002994. https://doi.org/10.1029/2022MS002994

Champion, K., Lusch, B., Kutz, J. N., & Bruton, S. L. (2019). Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, *116*. https://doi.org/10.1073/pnas.1906995116

Chandrakar, K. K., Grabowski, W. W., Morrison, H., & Bryan, G. H. (2021). Impact of Entrainment Mixing and Turbulent Fluctuations on Droplet Size Distributions in a Cumulus Cloud: An Investigation Using Lagrangian Microphysics with a Subgrid-Scale Model. *Journal of the Atmospheric Sciences*, *78*(9), 2983–3005. https://doi.org/10.1175/JAS-D-20-0281.1

Diamond, M. S. (2023). Detection of large-scale cloud microphysical changes within a major shipping corridor after implementation of the International Maritime Organization 2020 fuel sulfur regulations. *Atmospheric Chemistry and Physics*, *23*(14), 8259–8269. https://doi.org/10.5194/acp-23-8259-2023

Fierce, L., & McGraw, R. L. (2017). Multivariate quadrature for representing cloud condensation nuclei activity of aerosol populations. *Journal of Geophysical*

Research: Atmospheres, *122*(18), 9867–9878. https://doi.org/10.1002/2016JD026335

Gettelman, A., Gagne, D. J., Chen, C.-C., Christensen, M. W., Lebo, Z. J., Morrison, H., & Gantos, G. (2021). Machine Learning the Warm Rain Process. *Journal of Advances in Modeling Earth Systems*, *13*(2), e2020MS002268.

Grabowski, W. W. (2019). Separating physical impacts from natural variability using piggybacking technique. *Advances in Geosciences*, *49*, 105–111. https://doi.org/10.5194/adgeo-49-105-2019

Hill, A. A., Lebo, Z. J., Andrejczuk, M., Arabas, S., Dziekan, P., Field, P., Gettelman, A., Hoffmann, F., Pawlowska, H., Onishi, R., & Vié, B. (2023). Toward a Numerical Benchmark for Warm Rain Processes. *Journal of the Atmospheric Sciences*, *80*(5), 1329–1359. https://doi.org/10.1175/JAS-D-21-0275.1

Jokulsdottir, T., & Archer, D. (2016). A stochastic, Lagrangian model of sinking biogenic aggregates in the ocean (SLAMS 1.0): Model formulation, validation and sensitivity. *Geoscientific Model Development*, *9*(4), 1455–1476. https://doi.org/10.5194/gmd-9-1455-2016

Lamb, K. D., Walqui, M. v.-L., Santos, S. P., & Morrison, H. (2023). *Reduced Order Modeling for Linearized Representations of Microphysical Process Rates* (preprint). Preprints. https://doi.org/10.22541/essoar.168995384.44033471/v1

Manshausen, P., Watson-Parris, D., Christensen, M. W., Jalkanen, J.-P., & Stier, P. (2022). Invisible ship tracks show large cloud sensitivity to aerosol [Number: 7930 Publisher: Nature Publishing Group]. *Nature*, *610*(7930), 101–106. https://doi.org/10.1038/s41586-022-05122-0

Marchisio, D. L., & Fox, R. O. (2005). Solution of population balance equations using the direct quadrature method of moments. *Journal of Aerosol Science*, *36*, 43–73. https://doi.org/10.1016/j.jaerosci.2004.07.009

McGraw, R. (1997). Description of Aerosol Dynamics by the Quadrature Method of Moments, *Aerosol Science and Technology*, *27*. https://doi.org/10.1080/02786829708965471

Morrison, H., Chandrakar, K. K., Shima, S. .-., Dziekan, P., & Grabowski, W. W. (2024). Impacts of stochastic coalescence variability on warm rain initiation using lagrangian microphysics in box and large-eddy simulations. *Journal Of The Atmospheric Sciences*. https://doi.org/10.1175/JAS-D-23-0132.1

Morrison, H., van Lier-Walqui, M., Kumjian, M. R., & Prat, O. P. (2019). A Bayesian Approach for Statistical–Physical Bulk Parameterization of Rain Microphysics. Part I: Scheme Description. *Journal of the Atmospheric Sciences*, *77*(3), 1019–1041.

Mueller, M. E., Blanquart, G., & Pitsch, H. (2009). Hybrid method of moments for modeling soot formation and growth. *Combustion and Flame*, *156*, 1143–1155. https://doi.org/10.1016/j.combustflame.2009.01.025

Scheffer, M., Baveco, J. M., DeAngelis, D. L., Rose, K. A., & van Nes, E. H. (1995). Super-individuals a simple solution for modelling large populations on an individual basis. *Ecological Modeling*, *80*, 161–170. https://doi.org/10.1016/0304-3800(94)00055-M

Shima, S.-I., Sato, Y., Hashimoto, A., & Misumi, R. (2020). Predicting the morphology of ice particles in deep convection using the super-droplet method: Development and evaluation of SCALE-SDM 0.2.5-2.2.0, -2.2.1, and -2.2.2. *Geoscientific Model Development*, *13*(9), 4107–4157. https://doi.org/10.5194/gmd-13-4107-2020

Thornton, J. A., Virts, K. S., Holzworth, R. H., & Mitchell, T. P. (2017). Lightning enhancement over major oceanic shipping lanes. *Geophysical Research Letters*, *44*(17), 9102–9111. https://doi.org/10.1002/2017GL074982

Zsom, A., & Dullemond, C. P. (2008). A representative particle approach to coagulation and fragmentation of dust aggregates and fluid droplets. *Astronomy & Astrophysics*, *489*, 931–941. https://doi.org/10.1051/0004-6361:200809921