



Design and First Applications of Cleo, a Flexible Superdroplet Model for Warm-Cloud Physics on High-Performance Computers



Clara J.A. Bayley

Hamburg 2026

Hinweis

Die Berichte zur Erdsystemforschung werden vom Max-Planck-Institut für Meteorologie in Hamburg in unregelmäßiger Abfolge herausgegeben.

Sie enthalten wissenschaftliche und technische Beiträge, inklusive Dissertationen.

Die Beiträge geben nicht notwendigerweise die Auffassung des Instituts wieder.

Die "Berichte zur Erdsystemforschung" führen die vorherigen Reihen "Reports" und "Examensarbeiten" weiter.

Anschrift / Address

Max-Planck-Institut für Meteorologie
Bundesstrasse 53
20146 Hamburg
Deutschland

Tel./Phone: +49 (0)40 4 11 73 - 0

Fax: +49 (0)40 4 11 73 - 298

name.surname@mpimet.mpg.de

www.mpimet.mpg.de

Notice

The Reports on Earth System Science are published by the Max Planck Institute for Meteorology in Hamburg. They appear in irregular intervals.

They contain scientific and technical contributions, including PhD theses.

The Reports do not necessarily reflect the opinion of the Institute.

The "Reports on Earth System Science" continue the former "Reports" and "Examensarbeiten" of the Max Planck Institute.

Layout

Bettina Diallo and Norbert P. Noreiks
Communication

Copyright

Photos below: ©MPI-M

*Photos on the back from left to right:
Christian Klepp, Jochem Marotzke,
Christian Klepp, Clotilde Dubois,
Christian Klepp, Katsumasa Tanaka*



Design and First Applications of Cleo,
a Flexible Superdroplet Model for Warm-Cloud Physics
on High-Performance Computers



Clara J.A. Bayley

Hamburg 2026

Clara J.A. Bayley

aus London, Vereinigtes Königreich

Max-Planck-Institut für Meteorologie

The International Max Planck Research School on Earth System Modelling
(IMPRS-ESM)

Bundesstrasse 53

20146 Hamburg

Tag der Disputation: 31. März 2026

Folgende Gutachter empfehlen die Annahme der Dissertation:

Prof. Dr. Bjorn Stevens

Prof. Dr. Raphaela Vogel

Vorsitzender des Promotionsausschusses:

Prof. Dr. Dirk Notz

Dekan der MIN-Fakultät:

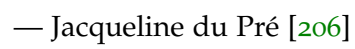
Prof. Dr.-Ing. Norbert Ritter

Cover photo: The cover photo was taken by this author on 12th September 2024 at sunset on Dover Beach, Barbados during the Organized Convection and EarthCARE Studies over the Tropical Atlantic

Berichte zur Erdsystemforschung / Max-Planck-Institut für Meteorologie
Reports on Earth System Science / Max Planck Institute for Meteorology

304
2026

ISSN 1614-1199 - doi: 10.17617/2.3717814



whom I wouldn't be whole.

ABSTRACT

Eulerian bulk and bin models for cloud microphysics hinder our understanding of warm tropical clouds in ways the Super-Droplet Method (SDM) can overcome. Microphysics-related uncertainties across global models and Large Eddy Simulations (LES), as well as discrepancies from observations, afflict precipitation patterns, extremes, radiation budgets, and cloud feedbacks. They arise not only from insufficient knowledge of cloud microphysics, but also modelling choices which bulk and bin models make to represent hydrometeors and microphysical processes. SDM, being more fundamental and Lagrangian, offers many advantages including eliminating these choices and facilitating exploration of knowledge gaps. However, the computation costs of current SDMs preclude their application to large enough domains for elucidating the microphysics-mediated coupling between shallow cumulus organisation and precipitation.

We create Cleo, a novel C++ SDM implementation for high-performance computers, seeking to make larger domain SDM computationally feasible. Cleo's memory layout is designed to minimise memory allocations and optimise access patterns. We use Kokkos (shared memory) and Message Passing Interface (distributed memory) parallelism, and couple to fluid-dynamical drivers via Yet Another Coupler, which allows total independence of their grids and domain decompositions. Thus Cleo enables efficient resource allocation as well as portability across different high-performance computer architectures. Cleo shows ideal superdroplet-scaling and reasonable strong-scaling and is ready for its final stage of development for large-domain LES: higher efficiency particle transport across distributed memory.

Equally important as the computational performance is that Cleo flexibly models warm-cloud microphysics. In up to 3-D simulations Cleo can represent all the principal microphysical processes driving warm-clouds; and Cleo's design enables users to easily compile an arbitrary number and combination of different processes and data outputs, as well as different options for uncertain details of microphysics, e. g. terminal velocity parametrisations and collision kernels. We offer two ways to model droplet collisions: the canonical SDM algorithm and an extended version which includes our novel representation of collisional breakup. Cleo is therefore a fully-fledged SDM, whose flexibility facilitates detailed study of warm-cloud microphysics.

We apply Cleo to constrain uncertainty in precipitation relevant to shallow cumulus organisation. First, we examine the sensitivity of precipitation in the Kinematic Driver (KiD) framework to particular modelling choices in SDM. We find less sensitivity to the choices of superdroplet initialisation, collision efficiencies, and the number of superdroplets than previous studies supposed. Second, we constrain the amount of and controls on sub-cloud rain evaporation from shallow cumuli. In agreement with previous studies, microphysics rather than the thermodynamic environment predominantly determines such rain evaporation. Furthermore, droplet collisions have a negligible impact on rain evaporation whereas the ventilation effect can increase it by up to a factor of 10. Both studies advance warm-cloud precipitation modelling and encourage using Cleo to better understand shallow cumulus organisation.

This dissertation constitutes substantial progress for modelling warm-clouds through the development and application of a novel high-performance SDM. It opens the door to deciphering the role of microphysics in shallow cumulus organisation from a fundamentally different, process-based perspective. Beyond warm-clouds, Cleo's design makes it easily extendable to other particle-based components of our climate system, to clouds on exoplanets, and to help bridge the gap between field observations and cloud microphysics modelling.

ZUSAMMENFASSUNG

Eulerian Bulk- und Bin-Modelle für die Wolkenmikrophysik behindern unser Verständnis warmer tropischer Wolken in gewisser Weise ein, die mit der Super-Droplet-Methode (SDM) überwunden werden kann. Unsicherheiten in Bezug auf die Mikrophysik in globalen Modellen und Large-Eddy-Simulationen (LES) sowie Abweichungen von Beobachtungen beeinträchtigen Niederschlagsmuster, Extreme, Strahlungsbilanzen und Wolkenrückkopplungen. Sie entstehen nicht nur aus unzureichendem Wissen über Wolkenmikrophysik, sondern auch aus Modellierungsentscheidungen, die Bulk- und Bin-Modelle treffen, um Hydrometeore und mikrophysikalische Prozesse darzustellen. SDM, die grundlegender ist und eine Lagrange-Perspektive einnimmt, bietet viele Vorteile, darunter die Eliminierung dieser Modellierungsentscheidungen und die Erleichterung der Erforschung von Wissenslücken. Die Rechenkosten aktueller SDM-Implementierungen schließen jedoch ihre Anwendung auf ausreichend große Domänen aus, um die mikrophysikalisch vermittelte Kopplung zwischen der Organisation flacher Cumuluswolken und Niederschlägen aufzuklären.

Wir entwickeln Cleo, eine neuartige C++-SDM-Implementierung für Hochleistungscomputer, mit dem Ziel, die Berechnung größerer Domänen mit SDM rechnerisch zu ermöglichen. Cleos Speicherlayout ist darauf ausgelegt, Speicherzuweisungen zu minimieren und Zugriffsmuster zu optimieren. Wir verwenden Kokkos-Parallelität (gemeinsamer Speicher) und Message Passing Interface-Parallelität (verteilter Speicher) und koppeln an fluid-dynamische Treiber über Yet Another Coupler (YAC), was völlige Unabhängigkeit von deren Gittern und Domänenzerlegungen ermöglicht. Somit ermöglicht Cleo effiziente Ressourcenzuweisung sowie Portabilität über verschiedene Hochleistungscomputerarchitekturen hinweg. Cleo zeigt eine ideale Superdroplet-Skalierung, eine angemessene Strong-Skalierung und ist bereit für die letzte Entwicklungsphase für LES mit großen Domänen: einen effizienteren Partikeltransport über verteilten Speicher.

Genauso wichtig wie die Rechenleistung ist, dass Cleo die Mikrophysik von Warmwolken flexibel modelliert. In bis zu 3D-Simulationen kann Cleo alle wichtigen mikrophysikalischen Prozesse darstellen, die Warmwolken antreiben; und Cleos Design ermöglicht es Benutzern, auf einfache Weise eine beliebige Anzahl und Kombination verschiedener Prozesse und Datenausgaben sowie verschiedene Optionen für ungewisse Details der Mikrophysik, z. B. Parametrisierungen der Endgeschwindigkeit und Kollisionskernel, zusammenzustellen. Wir bieten zwei Möglichkeiten zur Modellierung von Tröpfchenkollisionen: den kanonischen SDM-Algorithmus und eine erweiterte Version, die unsere neuartige Darstellung des kollisionsbedingten Aufbrechens einschließt. Cleo ist somit eine vollwertige SDM, deren Flexibilität detaillierte Studien zur Mikrophysik von Warmwolken erleichtert.

Wir wenden Cleo an, um die Unsicherheit bei Niederschlägen zu begrenzen, die für die Organisation flacher Cumuluswolken relevant sind. Erstens untersuchen wir die Empfindlichkeit des Niederschlags im Kinematic-Driver (KiD)-Framework gegenüber bestimmten Modellierungsentscheidungen in SDM. Wir stellen fest, dass die Empfindlichkeit gegenüber den Entscheidungen zur Initialisierung von Superdroplets, zur Kollisionseffizienz und zur Anzahl der Superdroplets geringer ist als frühere Studien vermuten ließen. Zweitens begrenzen wir die Menge und die Steuerungsfaktoren der Verdunstung von Regen unter flachen Cumuluswolken. In Übereinstimmung mit früheren Studien bestimmt eher die Mikrophysik als die

thermodynamische Umgebung diese Regenverdunstung. Darüber hinaus haben Tröpfchenkollisionen einen vernachlässigbaren Einfluss auf die Regenverdunstung, während der Ventilatoneffekt diese um bis zu zehnmal erhöhen kann. Beide Studien bringen die Modellierung von Warmwolken-Niederschlägen voran und fördern die Verwendung von Cleo, um die Organisation flacher Cumuluswolken besser zu verstehen.

Diese Arbeit stellt einen maßgeblichen Fortschritt für die Modellierung von Warmwolken durch die Entwicklung und Anwendung einer neuartigen Hochleistungs-SDM dar. Sie bereitet den Weg dafür, die Rolle der Mikrophysik in der Organisation flacher Cumuluswolken aus einer grundlegend anderen, prozessbasierten Perspektive zu entschlüsseln. Über Warmwolken hinaus lässt sich Cleos Design leicht ausweiten: auf andere partikelbasierte Komponenten unseres Klimasystems; auf Wolken auf Exoplaneten; und um die Lücke zwischen Feldbeobachtungen und der Modellierung von Wolkenmikrophysik zu überbrücken.

PUBLICATIONS

I am first-author of the following under-review publications which are part of this dissertation and included in the appendices:

C. J. A. Bayley, T. Kölling, A. K. Naumann, R. Vogel and B. Stevens, 'CLEO: The Fundamental Design for High Computational Performance of a New Superdroplet Model', *EGUsphere* **2025**, 1–33 (2025)

C. J. A. Bayley, A. K. Naumann, F. Poydenot, R. Vogel, B. Stevens and S.-I. Shima, 'CLEO: The Numerical Methods of a New Superdroplet Model including a Droplet Breakup Algorithm', *EGUsphere* **2025**, 1–26 (2025)

C. J. A. Bayley* and N. Niebaum*, F. Poydenot, A. K. Naumann, M. Sarkar and R. Vogel, 'Constraining Rain Evaporation from Shallow-Clouds in the Trades using an Observation-Based Superdroplet Model', *EGUsphere* **2025**, 1–30 (2025) *These authors contributed equally to this work.

I am first-author of the following manuscript in preparation for publication which is part of this dissertation and included in the appendices:

C. J. A. Bayley, E. Ware and S. Arabas, 'SDM as a Numerical Benchmark for Warm-Rain Processes? (in preparation)', (2025)

I am the main developer of the CLEO and microphysics test-cases code repositories described in this dissertation:

C. J. A. Bayley et al., *CLEO*: github.com/yoctoyotta1024/CLEO, (last accessed Jan. 2026)

C. J. A. Bayley, J. Behrens and G. Mania, *microphysics_testcases*: github.com/yoctoyotta1024/microphysics_testcases, (last accessed Jan. 2026)

I also contributed as second-author to a further publication during my PhD:

Segura, H. and Bayley, C. et al., 'A Single Tropical Rainbelt in Global Storm-Resolving Models: The Role of Surface Heat Fluxes Over the Warm Pool', en, *Journal of Advances in Modeling Earth Systems* **17**, 10.1029/2024MS004897 (2025)

*You can't stay in your corner of the Forest
waiting for others to come to you.
You have to go to them sometimes.*

— A. A. Milne [175]

ACKNOWLEDGMENTS

I can't quite find the words to show my appreciation for the Bayleys, to whom this dissertation is dedicated. Above all thank you Tiggy; thank you Dad; thank you Mum; and of course also thank you Susan and Ben. Thank you for the (unconditional) love you have given me, and for all your encouragement and enthusiasm for what I do — even though it means I live in another country. I am truly grateful for and proud of my family. Words don't express this enough.

Beyond the Bayleys, first and foremost thank you Anthony (Rev. B.) because I absolutely am a better person for knowing you. Thank you for being all that you are, and thank you for helping me get through the rough times during Alleyns so I can enjoy good days like today. Karim and Richard, your lessons also shaped me and helped me more than most. Thank you both.

Thank you to my wonderful friends who supported me through school and also afterwards. Natalie, Minka, Alice, and Louis I feel especially grateful for all the times you supported me and believed in me, especially when I couldn't do that for myself. Hazel, you are the definition of a *true* friend, thank you for being mine. Lawrence, thank you for so many hilarious and also brilliant shared supervisions, I feel like we got through Cambridge together. Alana, thank you for all the laughing and your energy; the way you filled a room still inspires me. Navi, Lily, Ralph, Felix, Jack, the Sams, and my other closest friends who I've missed in this list, thank you also.

Thank you also to my newer friends for brightening my time in Hamburg. Moritz and Helene I am massively grateful for you two as honest, caring and dedicated friends. (On top of that but really much less important, thank you and my parents again for your feedback on my unifying essay.) To name a few more, thank you Quan, Arim, Hernan, Florian, Angel, Hans, and Jairo for making me so happy here. Certain members of my football team as well, Börke, Schneck, Desi, and Kiki above all. Also Simon, I'm grateful you let me garden with you, even when I showed up late or was too busy. Lastly, thanks to the members of the DTC group, Jule and Theresa, from whom I learned so much in the first couple years of my PhD; and thank you Nils for being an exceptionally capable Master's student as well as very kind, it was a pleasure to work with you.

More broadly, thank you to the many hard-working people at MPI-M and DKRZ for making my PhD so fulfilling. Thank you to my very intelligent colleagues including Tobi, Sergey, Lukas, Wilton, and Moritz, and also to my other collaborators and co-authors. Thank you to all the members of MPI-M's administration, but especially Angela, Antje, Connie, and Michaela, who spent considerable time making things happen for me. Thanks also to all the other people who supported the NatESM sprints, ORCESTRA, and my many work-related travels. Further behind the scenes but still greatly appreciated, thanks to the DKRZ help-desk and the cleaners and maintenance staff at MPI-M.

For all the knowledge and experience you have shared with me, thank you to my teachers and non-PhD supervisors. Dr. de Silva and Mr. Jones you were especially important for getting me to pursue physics. Scott Annett thank you for genuinely caring about and helping me while I was at Cambridge. Matt, Steffi and Ziri, thank you for your time and teaching when I was a research student. Shin-ichiro, thank you for hosting me in your group in Japan; it was not just a great honour for me to learn from you, but also immensely beneficial.

Bjorn, Raphaela and Ann Kristin, I cannot express my gratitude to each of you enough. Thank you! You have given me so much of your time and energy, so much useful advice, so many opportunities, so much feedback on abstracts, posters, presentations, texts, etc. Most of all, I feel you have always given me proper attention, and have been conscientious of my academic development and more importantly my well-being. In short, I couldn't have wished for better supervisors.

Lastly, Annika. Thank you. Thank you for your patience, for accepting my stress levels and my low energy at the end of each long day; and thank you for the ways you love me. Just writing your name makes me smile.

CONTENTS

Unifying Essay	1
1 INTRODUCTION	3
1.1 The Demand for a Better Model for Cloud Microphysics	3
1.2 The Advent of Superdroplet Method Microphysics . . .	12
1.3 The Requirement for a New SDM	16
2 A NEW SUPERDROPLET MODEL CALLED CLEO	21
2.1 Creating a SDM for Large Domain Simulations	21
2.2 Model Design	22
2.3 Model Evaluation	29
2.4 Summary	34
3 USING CLEO TO CONSTRAIN UNCERTAINTIES IN WARM-CLOUDS	35
3.1 Constraining Inherent Model Uncertainty in SDM . . .	35
3.2 Constraining Rain Evaporation	42
4 CONCLUSION	49
4.1 Closing Thoughts	49
4.2 Outlook	52
Appendices	55
A CLEO: THE FUNDAMENTAL DESIGN	57
A.1 Abstract	58
A.2 Introduction	58
A.3 Memory Layout	60
A.4 Resource Allocation	64
A.5 Flexibility through Monoids	66
A.6 Time-stepping Cleo	71
A.7 Performance	72
A.8 Conclusions	78
A.9 Code availability	80
A.10 Code and Data availability	80
A.11 Competing interests	80
A.12 Acknowledgements	80
A.13 Large Figures	81
B CLEO: THE NUMERICAL METHODS	87
B.1 Abstract	88
B.2 Introduction	88
B.3 Condensation/Evaporation	90
B.4 Collisions	92
B.5 Motion	99
B.6 Validations	100
B.7 Conclusions	103

B.8	Code availability	104
B.9	Code and Data availability	104
B.10	Appendix	104
B.11	Competing interests	105
B.12	Acknowledgements	105
B.13	Large Figures	106
C	CONSTRAINING RAIN EVAPORATION FROM SHALLOW- CLOUDS	109
C.1	Abstract	110
C.2	Introduction	110
C.3	Data and Methods	112
C.4	Results and Discussion	116
C.5	Summary and Outlook	123
C.6	Code availability	126
C.7	Code and Data availability	126
C.8	Appendix	126
C.9	Competing interests	130
C.10	Acknowledgements	130
C.11	Large Figures	130
D	SDM AS A NUMERICAL BENCHMARK FOR WARM-RAIN PROCESSES?	133
D.1	Abstract	134
D.2	Introduction	134
D.3	KiD Setup	137
D.4	Results	141
D.5	Conclusion	145
D.6	Acknowledgements	147
D.7	Large Figures	147
	BIBLIOGRAPHY	149

ACRONYMS

Cleo	The novel implementation of SDM described in this dissertation, named after Cleopatra VII Thea Philopator as well as the anonymous mathematician
AoS	Array of Structures
BCO	Barbados Cloud Observatory
CCN	Cloud Condensation Nuclei
CPU	Central Processing Unit
DKRZ	Deutsche KlimaRechenZentrum
DNS	Direct Numerical Simulation
DSD	Droplet Size Distribution
EarthCARE	Earth Cloud, Aerosol and Radiation Explorer
EUREC ⁴ A	Elucidating the Role of Clouds-Circulation Coupling in Climate
GCM	Global Climate Model
GPU	Graphics Processing Unit
GSRMs	Global Storm Resolving Models
HALO	High Altitude Long range research aircraft
HAMP	HALO Microwave Package
HPC	High-Performance Computing
H ₂₀₂₃	Hill et al. [94]
ICON	ICOsahedral Nonhydrostatic (Model)
KiD	Kinematic Driver
LES	Large Eddy Simulations
LWP	Liquid Water Path
MCS	Mesoscale Convective System
MJO	Madden-Julian Oscillation
MPI	Message Passing Interface
NatESM	National Earth System Modeling (Project)
ORCESTRA	Organized Convection and EarthCARE Studies over the Tropical Atlantic
RICO	Rain In Cumulus over the Ocean
ROTOR	Rain and cloud Organization in the Trades using Observations and models

RWC	Rain Water Content
SDM	Super-Droplet Method; also known as Superdroplet Model
SoA	Structure of Arrays
specMACS	spectrometer of the Munich Aerosol Cloud Scanner
WALES	Water Vapor Lidar Experiment in Space
YAC	Yet Another Coupler

UNIFYING ESSAY

A search for truth seems to me to be full of pitfalls. We all have different understandings of what truth is ... I think a search for understanding is much more serviceable to humankind, and is a sufficiently ambitious goal of itself.

— Jocelyn Bell Burnell [30]

INTRODUCTION

*Perhaps the next generation will be talking
about the dynamics of water systems.*

— E. N. Lorenz [160]

1.1 THE DEMAND FOR A BETTER MODEL FOR CLOUD MICROPHYSICS

Water is one of the most phenomenal substances on Earth. Looking at the atmosphere alone, water naturally exists in three states of matter. It can convert between any of those states directly or indirectly. Water droplets and ice crystals suspended in the sky are just so right-sized and numerous as to scatter all wavelengths of visible light equally and appreciably, meaning clouds globally reflect nearly 30% of incoming solar radiation. Although water vapour rarely exceeds 3% of the atmosphere's composition, it is a potent greenhouse gas, responsible for at least 50% of the increase in the global greenhouse effect. The specific latent heat of vaporisation of water is an order of magnitude greater than that of any other common gas in the atmosphere, which makes water disproportionately powerful over convection, giving up enormous energy supplies where it condenses and taking it away again where it evaporates. The result is that clouds and precipitation are not at the mercy of the atmospheric dynamics, instead they play a starring role shaping their own story. In some ways perhaps it is no surprise that modelling water in the atmosphere remains one of the greatest challenges of 21st Century climate modelling.

1.1.1 *Emergent Problems Stemming from (Warm-)Cloud Microphysics*

1.1.1.1 *In Global Storm Resolving Models*

The 2020s are ushering in a new age of climate modelling, the age of Global Storm Resolving Models (GSRMs). Concomitant with the rapid rise in computing power, GSRMs can now be run globally without the need to parametrise convection and with grid-spacings less than 10 km [41, 74, 137, 210, 279, 280, 301, 309]. These scales are comparable to those we make observations at, meaning we can use global models and observations in unison more easily and transparently than ever before [198, 229, 252]. Moreover we are now modelling at scales relevant to society, making regional predictions of climate change globally [198]. From a modeller's perspective, the power of GSRMs comes

from removing parametrisations and letting the fundamental equations that we understand model convection instead.

Replacing parametrised with explicitly resolved convection has improved our representation and understanding of Earth’s climate. In contrast to conventional Global Climate Models (GCMs) which have struggled with convective parametrisations for decades, particularly when it comes to capturing shallow clouds, precipitation extremes, and a number of observed trends over The Tropics and Southern Ocean [46, 49, 64, 219, 240, 253, 304], GSRMs more accurately simulate many important features of the Earth’s atmosphere, including tropical cyclones [124], Mesoscale Convective Systems (MCSs) [55], precipitation (across a wide range of measures) [41, 163, 260, 262, 305], shallow cloud cover [41, 262], the structure of deep convection [41, 262], the diurnal cycle [41, 55, 190, 197, 225, 260, 262], tracer transport [229], and the atmosphere’s kinetic energy spectra [41, 229, 262]. These improvements have in some cases been directly attributed to the removal of convective parametrisations [e.g. 14, 129]. GSRMs have also led to a much greater appreciation for the importance of interactions between scales of convective systems, and especially the role of the mesoscale in The Tropics (specifically meso-beta, 20 – 200 km, [195]) [229, 252, 279], whose misrepresentation or omission in GCMs for example falsely predicts cloud cover changes under global warming [297]. The evidence is unequivocal, removing quasi-empirical, underdetermined parametrisations in favour of more fundamental equations is not just a matter of beauty, it is essential for better prediction and understanding of the climate system.

Despite advances across a range of metrics, GSRMs still fail to capture many aspects of tropical convection. Tropical convection is intensely important for the Earth’s climate, being responsible for much of the Earth’s precipitation, redistribution of energy, and modes of variability. However the infamous double Inter-Tropical Convergence Zone bias pervades GCMs and GSRMs alike, although recent work has rectified this to some extent in atmosphere-only simulations (Takasuka et al. [271], Segura, Bayley et al. 2025 [234]). Despite marked improvement compared to GCMs, GSRMs still suffer from the “too little, too often” drizzle bias [41] and sometimes they simulate too much liquid water in too small clouds due to an overabundance of “popcorn-like” convection [14, 41, 176]. Perhaps related, yet somewhat conflicting, is the too rapid transition from shallow to deep convection in GSRMs, which misses mid-level clouds [197, 277], as well as the difficulty GSRMs have transitioning from deep convection to MCSs [267], which may have implications for the Madden-Julian Oscillation (MJO). Whilst ever finer grid-spacing can improve the representation of The Tropics to some extent [125, 164, 200, 262], there is certainly more at play than just resolution.

Many of the biases [GSRMs](#) have representing tropical clouds appear to be related to uncertainties in cloud microphysics. By removing parametrised convection, cloud microphysics and subgrid-scale turbulence contain most of the remaining “tuning parameters” for [GSRMs](#). From this viewpoint, perhaps it is no surprise that microphysics is frequently adjusted to suit the characteristics of particular [GSRMs](#) to a particular purpose (e.g. the [MJO](#) [[220](#), [268](#), [272](#)], tropical cirrus [[66](#)], the vertical atmospheric temperature profile [[239](#)], ...). The sensitivity of [GSRMs](#) to seemingly innocuous and/or arbitrary features of different microphysics models is however, extreme. Simply changing between microphysics schemes can inadvertently alter the speed of precipitation formation [[265](#)], which has repercussions for the inversion height in the trades; precipitation efficiency; and the Top-Of-Atmosphere energy budget, changing outgoing radiation in some cases by several W m^{-2} [[140](#), [185](#)]. Setting aside the numerous ways in which relatively small changes to ice microphysics models can propagate into large differences between [GSRMs](#), even model parameters related to warm-rain can be highly consequential, as shown for example by the sensitivity of tropical precipitation extremes [[17](#)]. It is also still not entirely clear how modifying warm-rain affects shallow cloud biases in [GSRMs](#), and how that could influence the heat budget of the cloud layer as well as the larger-scale environment [[186](#)].

1.1.1.2 *In Large Eddy Simulations*

The deficiencies of Large Eddy Simulations ([LES](#)) in representing principal features of clouds are indisputably linked to inadequate representation of cloud microphysics. Just like their global cousins, [LES](#) struggle to reproduce observed cloud-type transitions and spatial patterns of clouds and precipitation. Moreover, for [LES](#) an extensive literature exists which ties these biases to problematic features of cloud microphysics models. Excessive cloud-top entrainment depleting stratocumulus has been attributed to a lack of cloud droplet sedimentation in microphysics schemes [[3](#)], and further errors affecting cloud liquid water arise from schemes failing to represent supersaturation [[130](#)], and/or their ability to create spurious supersaturations [[256](#)]. The inability of [LES](#) to match satellite observations has also been shown to be a direct result of too rapid warm-rain formation [[131](#)]. Repeatedly throughout the literature, large macrophysical biases analogous to these ones have been attributed to microphysics schemes [e.g. [2](#), [93](#), [232](#), [289](#)], or even specific details of them, for example too much size-sorting or too fast raindrops [[2](#), [39](#), [157](#), [172](#), [292](#), [300](#)]. Even more clearly than [GSRMs](#), [LES](#) are suffering from high sensitivity to seemingly innocuous and/or arbitrary features of their microphysics schemes. The attribution of discrepancies between [LES](#) and observations to microphysics occurs even more frequently when ice

microphysics is included [e. g. 62, 63, 174, 182, 290–292, 307]. These issues have persisted for decades [84, 130, 182, 214], and were so severe that Randall et al. [214] began their article with the line “clouds and climate: a problem that refuses to die”. Over 20 years since then and the problem is still very much alive.

GSRMs and LES are putting pressure on the need to better model cloud microphysics so that we can capture the larger-scale properties of warm-clouds. In light of the new age of GSRMs and the ever-growing body of evidence from LES, the problems emergent from cloud microphysics have never felt more acute. In the following section I therefore explore where these problems originate from, in order to understand how an alternative model for cloud microphysics can eliminate (or at least alleviate) many of the issues arising from conventional cloud microphysics models. The alternative, fundamentally different model is the Super-Droplet Method; also known as Super-droplet Model (SDM).

The rest of this dissertation, and indeed much of the previous section, focuses on warm-cloud microphysics and its role for tropical shallow clouds, in particular trade-wind shallow cumuli. The reasons for this focus will be elucidated in Section 1.3.1. Although I have omitted them except where I thought it essential, at many points the role of ice microphysics, deeper and/or extra-tropical convection could equally well be discussed. Needless to say, the problems with and emergent from conventional models for cloud microphysics, as well as the capabilities of SDM, transcend warm tropical clouds.

1.1.2 The Fundamental Problems with (Warm-)Cloud Microphysics

1.1.2.1 The Root Cause

The principal equation which describes hydrometeors (liquid and solid water particles in the atmosphere) is the kinetic microphysics equation,

$$\frac{\partial f}{\partial t} + \mathbf{u} \cdot \nabla f - \frac{1}{\rho} \frac{\partial(\rho V f)}{\partial z} = \left(\frac{\partial f}{\partial t} \right)_{\text{diff}} + \sum_i^N P_i, \quad (1.1)$$

where $f \equiv f(\mathbf{x}, t, \mathbf{q})$ is the probability distribution that describes the hydrometeor population as a function of the location in physical space, $\mathbf{x}$, time, t , and hydrometeor attributes, $\mathbf{q}$ (size, shape, density, etc.); $\mathbf{u}$ is the advection velocity (i. e. wind); $V \equiv V(\mathbf{x}, t, \mathbf{q})$ is the hydrometeors' fall speed; $\rho \equiv \rho(\mathbf{x}, t)$ is the air density; $\left(\frac{\partial f}{\partial t} \right)_{\text{diff}}$ is diffusion of hydrometeors in physical space; and each P_i is a microphysical process (condensation/evaporation, deposition, fragmentation, etc.)

[182]. This equation looks remarkably similar to the Boltzmann equation,

$$\frac{\partial f}{\partial t} + \mathbf{u} \cdot \nabla f + \frac{\mathbf{F}}{m} \cdot \frac{\partial f}{\partial \mathbf{u}} = \left(\frac{\partial f}{\partial t} \right)_{\text{coll}}, \quad (1.2)$$

where $f \equiv f(\mathbf{x}, t, \mathbf{v})$ is the probability distribution that describes a gas particle population as a function of the location in physical space, $\mathbf{x}$, time, t , and particles' velocity, $\mathbf{v}$; $\mathbf{u}$ is the advection velocity; $\mathbf{F} \equiv \mathbf{F}(\mathbf{x}, t)$ is the net external force acted upon the particles; m is the particles' mass; $\left(\frac{\partial f}{\partial t} \right)_{\text{coll}}$ is collisions between particles; and $\frac{\partial f}{\partial \mathbf{u}} \equiv \nabla_{\mathbf{u}} f$. The only fundamental difference between Equation (1.1) and Equation (1.2) is that the distribution function in Equation (1.1), $f(\mathbf{x}, t, \mathbf{q})$, is a function of hydrometeor attributes, whereas in Equation (1.2), $f(\mathbf{x}, t, \mathbf{v})$ only depends on the gas particles' velocity.

The Boltzmann equation describes individual gas particles in the atmosphere but can be approximated for their bulk properties by the Navier-Stokes equations. The Boltzmann equation well describes the evolution of a population of gas particles, meaning it is the equivalent for gases in the atmosphere to the kinetic microphysics equation for hydrometeors. Fortunately, rather than solve the Boltzmann equation, climate models only need to solve an approximation to it, the Navier-Stokes equations (at least for pressures greater than $\mathcal{O}(1 \text{ Pa})$, up-to about 500 km on Earth [182, 288]). The Navier-Stokes equations describe the bulk properties of the fluid comprised of the gas particle population; namely the fluid's velocity, temperature, pressure and density, and so rather than model each individual gas particle in the atmosphere, we can more simply model their bulk statistics. Similarly, if there was an equivalent approximation for the kinetic microphysics equation, we would no longer need to model hydrometeor populations explicitly in climate models.

All the problems with cloud microphysics in climate models stem from the fact that we do not (yet) know the Navier-Stokes-like equations for describing hydrometeors. The approximation of the Boltzmann equation to the Navier-Stokes equations can be derived in several ways [138, 288], but key to any derivation is that the particle distribution is “stationary”, in other words, the same everywhere [31, 138, 288]. This is true for gas particles in the atmosphere, which are well-described universally by the Maxwell-Boltzmann distribution. However, a universal distribution describing hydrometeors has certainly not been found¹ and therefore the assumption made to ap-

¹ The immediate question raised here might be why is there no universal particle distribution for hydrometeors? But perhaps a better question would be why does the Maxwell-Boltzmann distribution universally describe gas particles in the atmosphere? A universal distribution exists because the timescale for equilibration of gas particles (through collisions) is much smaller than the timescales for the motion of interest, and thus gas particles everywhere follow their equilibrium distribution. Why the Maxwell-Boltzmann distribution is the equilibrium distribution of gas particles

proximate the Boltzmann equation by the Navier–Stokes equations is not valid for the kinetic microphysics equation [182]. Thus to obtain the bulk behaviour of clouds we have to attempt to model the entire hydrometeor population of the atmosphere. This is where the challenge of cloud microphysics modelling begins.

The attempt to model every hydrometeor in the atmosphere introduces both knowledge uncertainty and inherent model uncertainty into our climate models. Knowledge uncertainty is epistemic uncertainty. It can be defined as uncertainty due to a lack of understanding of the hydrometeor population, which we could resolve with more data. From Equation (1.1) we can see this in terms of poorly constrained yet important P_i or V , and/or insufficient understanding of the relevant $\mathbf{q}$ for describing the hydrometeor population. Inherent model uncertainty is uncertainty bound to *the way* we model cloud microphysics, irrespective of gaps in our knowledge. Given Equation (1.1), we may choose to solve for f by modelling individual particles, discretising f into discrete bins of x and $\mathbf{q}$, or modelling some number of moments of f (integrals over $\mathbf{q}$). No matter what we decide, we are inevitably forced to make arbitrary modelling choices, and when the solution of f is highly sensitive to these choices, our simulations suffer from inherent model uncertainty. Regarding the problems emergent from cloud microphysics discussed in Section 1.1.1, both types of uncertainty are their source.

1.1.2.2 Knowledge Uncertainties

Substantial gaps remain in our understanding of cloud microphysics. Whilst those related to ice microphysics are arguably more concerning, with large uncertainties remaining in our understanding of both ice properties and the microphysical processes that affect them [182], nevertheless our knowledge of warm-cloud microphysics is also insufficient. Some measurements are still poorly constrained regarding Cloud Condensation Nuclei (CCN) and aerosol-cloud interactions [88, 147], but more serious knowledge gaps remain in our understanding of turbulence-microphysics interactions, which puts conventional models in a state of “dire need of improvement” [157]. Turbulent mixing in warm-clouds, the entrainment of dry air, and the effect of turbulence on droplet collisions and condensational growth are all relatively poorly constrained processes which may have serious implications for warm-cloud development as well as precipitation formation and efficiency [54, 84].

Arguably the largest uncertainties that persist in warm-cloud microphysics are related to droplet collisions and in particular collisional breakup [182]. LES are known to be sensitive to their represent-

is another matter... one could derive this from the Hamiltonian or Lagrangian of the particles, or from thinking very hard about Liouville’s theorem, or descending into quantum mechanics... Maybe here just accept the beauty of nature.

ation of droplet collisions and even models which ignore collisional breakup have considerable differences quantifying the probabilities with which collisions occur and then lead to droplet coalescence, exacerbated by turbulence-microphysics uncertainty [54]. On top of this, a lack of both measurements and theory describing the probability of collisional breakup and fragment size distributions make breakup extremely challenging to model [182]. What few laboratory measurements we do have show its importance for affecting Droplet Size Distributions (DSDs) [161, 162, 167, 169, 270], corroborated by Direct Numerical Simulations (DNSs) [29, 230], and changes to the DSD in this way can have significant macrophysical repercussions by affecting cloud organisation [178, 180, 203, 235, 236, 265], as will be explained in Section 1.3.1.2. Scarce in-field observations also seem to hint that breakup frequently occurs in heavy precipitation [275]. Thus large deficiencies in our understanding of collisions, and especially breakup, could be significant sources of uncertainty in climate models.

1.1.2.3 *Inherent Model Uncertainties*

Since the 1950s, there have been two types of conventional models for cloud microphysics. Bulk schemes, also known as moment schemes, are the workhorse of modern day climate models. They divide hydrometeors into different categories (cloud water, rain water, ice, snow, ...) and predict an arbitrary number of moments of the size distribution for each category in each grid-box (cell) of the simulated domain. In terms of Equation (1.1), they divide f into distinct categories with pre-defined attributes except for size/mass (i.e. fixed density, shape etc.) and then solve for integrals over the droplet size/mass distribution (DSD) at discrete x . Bulk schemes are desirable for their low computational cost and therefore they usually only predict the 1st moment of the distribution, the total mass of each hydrometeor category; but some schemes additionally predict the 0th moment, the droplet number, to be less prescriptive about the processes controlling it. More recently people have also begun advocating for three-moment schemes [173, 179], and arbitrary moment schemes [107]. (Spectral) Bin schemes attempt to model the hydrometeor size distribution more accurately than their bulk counterparts at the price of higher computational cost. They divide hydrometeors into a liquid and several ice categories but rather than model moments, they model the evolution of a discretised version of the DSD for each category. In other words, rather than solve for integrals over size/mass, they solve f for many predefined bins of it instead. Both bulk and bin schemes suffer from many serious inherent model uncertainties [84, 182].

Bulk microphysics schemes introduce large uncertainty because they make numerous arbitrary and ill-founded modelling choices to represent hydrometeors. These choices are ubiquitous in the liquid

phase and further exacerbated when the ice-phase is modelled too. The first assumption, that hydrometeors can be divided into distinct categories, is poorly defined because a significant amount of hydrometeors exist in the grey-zone between categories, and becomes problematic because the conversion between categories is unphysical and creates artificial jumps in fall velocities [84, 157]. Choosing different category definitions therefore has large macrophysical implications [e.g. 93, 104, 106], and whilst reducing the number of hydrometeor categories eliminates some of these issues, it introduces new ones by creating more complicated parametrisations for microphysical processes, with more free parameters [179]. Even given some choice for hydrometeor categories, bulk schemes must then make assumptions about how to relate moments to a specific DSD. However, an optimal set of moments to decide the DSD does not exist and the chosen functional form of the DSD is often inappropriate [109, 110, 188] (as of course could already be foreseen by the fact there is no universal distribution for describing hydrometeors). As a consequence, not just hydrometeor categories, but also loosely constrained parameters defining the chosen DSDs are a source of large uncertainty in any particular bulk scheme [e.g. 65, 108, 111, 172, 292].

As well as through their representation of hydrometeors, bulk schemes introduce uncertainty through their representation of microphysical processes. Whilst in reality the relation of the DSD to microphysical processes rates is highly non-linear and sensitive to the right-hand tail of the DSD, this is frequently lost by bulk scheme parametrisations [90, 130], a problem exacerbated by the fact that radar observations are also especially sensitive to this tail, as well as particle shapes and composition which are even more crudely represented by bulk schemes [187]. Whilst increasing moments gives more freedom to the form of the DSD and this can help to represent its tail, it also introduces new problems because the advection of different moments is intrinsically inconsistent [182]. Furthermore, parametrisations are typically so convoluted that sensitivity to different processes is undecipherable [207]. This enables a “dilemma of growth” [87], whereby the increasing complexity of the models jeopardises rather than advances understanding. For all these reasons it is very difficult to constrain parametrisations for microphysical processes in bulk schemes directly from observations and therefore rates for supposedly the same process can differ by orders of magnitude across schemes [130], which contributes yet another source of inherent uncertainty in bulk schemes. Finally, bulk schemes are often tuned to better represent particular meteorological conditions, but this is at unknown detriment to other conditions and is not robust across schemes [72, 130, 188]. Hence in bulk schemes the representation of both hydrometeors and microphysical processes are sources of inherent model uncertainty,

from which most of the issues attributed to microphysics in Section 1.1.1 emerge.

Since bin schemes more explicitly represent hydrometeor size distributions, they do not suffer from the same sources of inherent uncertainty as bulk schemes, however bin schemes are still profoundly problematic. Bin schemes do not have to make assumptions about the shape of the DSD and, for liquid clouds, they also do not suffer from the artificial categorisation of hydrometeors, i. e. they do not distinguish between cloud and rain water. Naturally, this also allows microphysical processes to be depicted more clearly. However, the first source of inherent uncertainty arises from numerical diffusion in bin schemes [84, 181, 254]. Numerical diffusion in physical space enhances entrainment of dry air which dilutes cloud-edges unphysically [79, 117, 171], and numerical diffusion in droplet-size space creates unphysical DSD broadening which confounds efforts to understand physical mechanisms for DSD broadening such as eddy-hopping [84, 100]. Numerical diffusion appears to be unavoidable despite decades of work and it still significantly deteriorates the quality of bin schemes in comparison with radar retrievals [145, 146].

Aside from numerical diffusion, the fundamental basis of bin schemes causes uncertainty. The structure of bin schemes means they suffer from the “curse of dimensionality” whereby, for a scheme with N bins and d different attributes describing a hydrometeor type, the computational cost increases by N^d . Beyond $d \approx 2$ bin schemes become computationally intractable, which restricts the representation of hydrometeors to their DSDs. This leads to crude modelling assumptions, and hence uncertainty, in the roles of aerosols and chemistry in warm-clouds, in spite of their accepted importance [84]. Perhaps less well known is that bin schemes are also fundamentally limited by the assumptions of the Smoluchowski equation used to describe droplet collisions [84, 182]. This is a mean-field equation and therefore assumes that the locations of differently sized droplets are uncorrelated [73], however this is only variably true [56, 182], and prevents bin schemes from modelling “lucky drops” which grow exceptionally rapidly through more frequent than average collisions and which may be crucial for precipitation onset [56, 135, 182, 273]. Thus uncertainty in warm-clouds also arises from the representation of droplet collisions in bin schemes. Finally, whilst microphysical process parametrisations in bin schemes are clearer than bulk schemes, they are by no means transparent [130], again confounding attempts to constrain parametrisations with direct comparison to observations. There was originally the expectation that bin schemes could be used as the numerical benchmark by which we could evaluate the fidelity of bulk schemes, but as a result of all these inherent uncertainties, there is as much disagreement amongst them as there is amongst bulk schemes

[94, 103, 289]. An alternative, fundamentally different approach to cloud microphysics is required.

1.2 THE ADVENT OF SUPERDROPLET METHOD MICROPHYSICS

1.2.1 *The Benefits of SDM*

1.2.1.1 *A Lagrangian Particle Model*

Lagrangian models for cloud microphysics eliminate or diminish many of the problematic features of conventional microphysics models and thus also their associated uncertainties. Bulk and bin cloud microphysics schemes are Eulerian, meaning they model the properties of the hydrometeor population, f in Equation (1.1) (or more simply the DSD), at fixed locations in space. In contrast Lagrangian microphysics schemes model the properties of the hydrometeors along their trajectories through space, meaning f at any location is diagnosed from the modelled hydrometeors afterwards. Many of the benefits of Lagrangian models arise from this fundamental difference in perspective. Whereas in bulk or bin models solving for f directly requires that the particle population is a priori defined by certain x and q , in Lagrangian models f is emergent from the particles and therefore there is no need for such definitions; every hydrometeor can have any x or q . Lagrangian models therefore eliminate all uncertainties arising from defining hydrometeor categories and assuming a functional form for the DSD, and furthermore they are intrinsically non-diffusive [84, 182].

Arguably the greatest strength of Lagrangian microphysics however, is their more fundamental representation of the underlying microphysics. Unlike bin and bulk schemes, which must parametrise how f (or the DSD) evolves as a result of microphysical processes, Lagrangian models represent the microphysical processes acting on each hydrometeor explicitly, replacing arbitrary model parameters with solving the fundamental equations which we believe govern hydrometeor motion, growth, and decline. The parallels with GSRMs are immediate. In the same way GSRMs improved our representation and understanding of Earth's climate by removing quasi-empirical, under-determined parametrisations in favour of fundamental equations, so too can Lagrangian models for cloud microphysics.

1.2.1.2 *The Collision-Coalescence Algorithm*

The crippling stumbling block for Lagrangian cloud microphysics was how to model collision-coalescence, until the late 2000s. Collision-coalescence is the process in which two droplets collide and subsequently merge together. The first Lagrangian cloud microphysics models were already being developed in the 1940s [102], and already

their numerous advantages compared to conventional models were being highlighted in the 1970s [47, 141]. However, by the early 2000s Lagrangian models were still struggling to represent collision-coalescence [113, 246], until a series of coincidental breakthroughs occurred demonstrating Lagrangian cloud microphysics with collision-coalescence in LES [9, 10, 215, 241, 249].

Of all the collision-coalescence algorithms, the one in SDM, introduced by Shima et al. [241], stands out for its advantageous properties of stochasticity, convergence, and computational tractability. In SDM, superdroplets are simulated particles which represent the entire real droplet population. Superdroplets have the same attributes as real droplets as well as an extra one, the multiplicity, which indicates how many real droplets each superdroplet represents. They are governed by the same fundamental equations as real droplets, except for collisions which are enacted stochastically with a Monte-Carlo scheme. This makes SDM a particle-in-cell (or grid-box) method rather than a purely particle method. By construction, the expected number of collisions represented by two superdroplets is equal to the expected number of real droplet collisions, but the variance is greater by a factor proportional to the higher multiplicity superdroplet [241].

Since the collision algorithm is stochastic and makes no mean-field approximation, it works even in the limit of small volumes where the Smoluchowski equation breaks down, and it can represent lucky drops [56]. On top of this the convergence property of SDM is particularly insightful. Since the modelled variance in collisions is the real variance scaled by the superdroplet multiplicities, as the superdroplet multiplicities decrease, i. e. the number of superdroplets increases, the collisions of superdroplets tend towards the real droplet description. SDM converges faster in this sense than any other Lagrangian microphysics scheme [177, 286, 287]. Furthermore, combined with the convergence of collision-coalescence, as the number of superdroplets increases, the “resolution” with which the real droplet population is sampled by superdroplets also increases, and therefore SDM converges towards what would be expected from a particle-by-particle DNS [182]. Such an appealing construction is in stark contrast to bin schemes and makes SDM a promising tool for a numerical benchmark for cloud microphysics [84, 94, 182]. The final appealing features of the SDM collision-coalescence algorithm are that it linearly samples superdroplets and conserves simulated-particle number. These have positive implications for the computational tractability of SDM.

1.2.1.3 Computational Aspects of SDM

SDM is well-suited to high-performance computers. Each superdroplet is independent of the others for all processes except collision-coalescence and collision-coalescence works well under linear-sampling, [177, 287], that is, at each time-step SDM enacts collisions between

randomly-generated pairs of superdroplets in a given volume rather than considering every possible pair of superdroplets. Thus all microphysical processes are trivially parallelisable and hence efficient on computers with a lot of parallel processing. Additionally, the conservation of simulated-particle number during collisions is beneficial for load balancing (the even distribution of tasks across computer processors) and memory organisation. Overall this means [SDM](#) performs well on high-performance computers, especially those with Graphics Processing Units ([GPU](#)s) [[11](#), [19](#), [57](#), [58](#), [152](#), [165](#)]. Based on various studies investigating [SDM](#)'s convergence, it is assumed that $\mathcal{O}(100)$ superdroplets per grid-box is sufficient to capture the mean properties of clouds [[177](#), [286](#), [287](#)]. Comparing this with the best performing superdroplet simulations to-date, [LES](#) using [SDM](#) can be run as fast as a two-moment (bulk) scheme, albeit with considerably more computations and a larger memory footprint [[165](#)].

[SDM](#) is also well-suited to trends in High-Performance Computing ([HPC](#)). [SDM](#) scales linearly in cost both with increasing number of superdroplets and increasing number of hydrometeor attributes (dimensions of $\mathbf{q}$), in stark contrast to bin schemes which scale quadratically with increasing number of bins and intractably with increasing attributes [[84](#), [241](#)] (Sec. [1.1.2.3](#)). In light of the fact that pressures on [HPC](#) are pushing advances in memory bandwidth and exascale systems with [GPU](#)s, looking ahead [SDM](#) looks very capable of harnessing the power of future computer architectures, to run more detailed simulations with increased numbers of superdroplets or hydrometeor attributes, or perhaps to even one day replace bulk schemes as the workhorse of climate models.

1.2.2 Practical Successes of [SDM](#) So Far

The theoretical advantages of [SDM](#) compared to conventional models for microphysics have already been made apparent in process-level understanding. Due to the explicit representation of condensation and the ability to include aerosol attributes easily, the role of [CCN](#) in aerosol-cloud interactions has been closely examined, for example to investigate how effectively giant [CCN](#) and lucky drops can trigger precipitation [[150](#), [154](#), [202](#)], the impact of marine cloud brightening [[97](#)], droplet activation and aerosol processing [[13](#), [83](#), [85](#), [98](#), [194](#)] — including the eradication of spurious supersaturations [[10](#), [12](#), [96](#), [181](#)]. Also due to the ease with which additional attributes can be modelled, isotopes, chemistry and ice properties have all been incorporated into [SDM](#) [[36](#), [119](#), [168](#), [238](#), [242](#), [249](#), [303](#)].

By eliminating numerical diffusion, [SDM](#) has been used to make marked progress on a wide range of topics related to entrainment and mixing [[1](#), [42](#), [82](#), [84](#), [96](#), [100](#), [155](#), [156](#)], droplet sedimentation and recirculation [[42](#), [188](#), [308](#)], and turbulence-microphysics inter-

actions [43, 86, 139, 148, 149, 151, 191]. To name just one example, Chandrakar et al. [43] were able to shed light on the “rain formation bottleneck” in a precipitating cumulus congestus. They showed the effect of turbulence on collision-coalescence explains the evolution of observed DSDs with height, and that neither giant-CCN nor superadiabatic droplet growth can explain the observations. Not only that, but precipitation onset can be 30 min sooner and its intensity over 7 times greater when accounting for turbulence in collision-coalescence, which naturally raises further questions about the repercussions had from neglecting such effects in conventional microphysics models. In this study as well as the others mentioned above, SDM has facilitated progress not just by reducing inherent model uncertainty, but also by making it easier to address knowledge gaps.

LES have also demonstrated the effectiveness of SDM for simulating various meteorological conditions. SDMs including ice microphysics have been productive, for example resolving large uncertainties related to contrail lifetimes [126]. Of particular relevance for this dissertation though, is the use of SDM to model shallow clouds. The numerous advantages of SDM, particularly with regard to numerical diffusion and CCN, have improved the representation of stratocumulus [8, 308], and enabled detailed studies of their closed to open cell transitions [45]. Better quality simulations of Pi Cloud Chamber experiments and of observations of non-precipitating shallow cumuli from the Rain In Cumulus over the Ocean (RICO) campaign have also been reported [77, 81]. Several other studies have also exploited SDM to study shallow cumulus with a common focus on how different microphysical processes (and resolution) can effect the timing of precipitation formation [12, 99, 187, 191, 227, 228], as also seen for cumulus congestus above. These studies frequently point out how removing numerical diffusion, avoiding assumptions on the DSD, and explicitly resolving microphysical processes have benefitted their simulations, in contrast to using bulk or bin schemes.

1.2.3 *Potential Drawbacks of SDM*

Despite the theoretical advantages and many successful applications of SDM to date, SDM comes with some of its own challenges. Despite eliminating many of the sources of inherent model uncertainty, SDM may not eradicate all inherent model uncertainty because it necessarily has to make some modelling choices of its own. SDM shows some sensitivity to the number of superdroplets per grid-box and the method chosen to initialise superdroplets [94, 177, 286, 287, 311], but the extent to which models using SDM suffer from inherent model uncertainty is not yet clear. Cautious modellers may also be alarmed by using a microphysics scheme in its relative infancy in contrast to

conventional bulk and bin schemes which have decades of expertise supporting them.

Most challenging of all however, is the computational expense of existing [SDMs](#). Whilst bulk schemes have many flaws, their low cost allows them to be used in numerical weather prediction daily all across the world and, at the other end of the spectrum, they are feasible in global climate simulations $\mathcal{O}(100\text{ yr})$. With careful tuning they are therefore able to fulfil many important services which [SDM](#) is currently far too expensive for. [SDM](#) necessarily has a much higher memory footprint than bulk schemes because the data stored per grid-box in [SDM](#) is much greater than the one or two values in a bulk scheme. Also in comparison to bulk schemes, the downside of having a much greater fidelity to fundamental microphysics is that calculations in [SDMs](#) are generally more complex. Therefore, even though we can exploit [SDM](#)'s trivial parallelisability on high-performance computers to make it as fast as bulk schemes, it still requires more compute. Furthermore, being a Lagrangian model [SDM](#) brings with it the challenges related to particle transport. Well-established methods for high computational performance achieve load balancing by relying on the workload of each computer processor staying relatively stable for a whole simulation. Conventional Eulerian models are therefore suitable for these methods whereas Lagrangian models, whose workload depends on the varying activity and number of particles, are not. Similarly, standard methods for communication between computer processors are not appropriate for transporting Lagrangian particles. Therefore methods which are less well-established and less straightforward are required to achieve load balancing and fast communication in [SDM](#). Notwithstanding the fact [SDM](#) is well-suited to [HPC](#) and its current trends, existing implementations of [SDM](#) are nevertheless expensive and achieving high performance is challenging.

1.3 THE REQUIREMENT FOR A NEW SDM

1.3.1 *Understanding the Organisation of Tropical Shallow Cumuli*

Given the growing evidence for the theoretical benefits and practical uses of [SDM](#), we decided to tackle the computational challenges of [SDM](#) so we could use it to study tropical shallow cumulus organisation. At the beginning of my PhD, not only were [GSRMs](#) starting to show the importance of the mesoscale, but also a wealth of studies were coming to fruition from EUcinating the Role of Clouds-Circulation Coupling in Climate ([EUREC⁴A](#)), the field campaign in the trade-wind region two years prior [35, 263]. In particular they realised the importance of shallow cumulus clouds' coupling to shallow mesoscale circulations [70, 71], and re-emphasised that shallow cumuli tended to organise into different mesoscale patterns [53, 116,

261], with different convective characteristics [33, 183, 211, 212, 231, 296]. Thus the importance of mesoscale tropical shallow cumulus organisation started to become apparent. At the same time, multiple lines of evidence pointed towards the role of warm-cloud microphysics for determining this organisation, and LES were conducted which displayed notable shortfalls when representing them and their precipitation [232, 294, 295, 299]. For example, compared to radar observations from EUREC⁴A, LES show too frequent light precipitation, an inability to capture heavy precipitation, and the “sugar” cloud pattern with too large cloud-base cloudiness [231]. We therefore sought to use SDM for warm-cloud microphysics to improve the representation of the organisation and precipitation of tropical shallow cumuli in LES; or at the very least better constrain the uncertainties in them which arise from microphysics. As will be explained in the next Chapter, that necessarily led us to create a new computational implementation of SDM; but first let us explore our motivation in more detail.

1.3.1.1 *The Importance of Tropical Shallow Cumulus Organisation*

The Earth’s energy budget is highly sensitive to changes in tropical shallow cumulus cloud-cover and this may be influenced by their mesoscale organisation. One of the major breakthroughs from EUREC⁴A was the realisation that mesoscale dynamics are absolutely critical for controlling cumulus cloud-cover [33, 296, 297]. Whilst this result discredited the possibility of large climate sensitivity stemming from an implausible feedback mechanism between tropical cumuli and tropospheric warming [the mixing-desiccation mechanism, 293, 297], it may still be possible for global warming to result in the expression of different mesoscale cloud patterns. Whether or not concomitant changes in cloud-cover would then have implications for Earth’s climate sensitivity is an open discussion. On the one hand, net cloud radiative effects appear to be a linear function of organisation [33], and therefore changes to the abundance of different forms of organisation could affect climate sensitivity [166]. On the other hand, a new theoretical framework for understanding mesoscale circulations shows that any changes to the short-wave cloud radiative effect as a result of mesoscale organisation is almost exactly offset by associated changes to the outgoing long-wave radiation from cloud-free regions [114]. Overall, although tropical shallow cumuli have a large impact on the Earth’s energy budget, the effect of their organisation on radiation remains a subject for debate, and therefore so too does the impact of potential changes to their organisation under global warming. Recent observations show shallow cumulus organisation has changed since the 1970s [59], and so the need has intensified to clarify what controls the expression of different mesoscale patterns and the influence of their changes on Earth’s energy budget.

Shallow cumulus organisation is clearly important for precipitation. Shallow clouds overall are responsible for over 70% of the precipitation in The Tropics [143], and 30% worldwide. Although total tropical precipitation is constrained by the Earth’s energy balance, the pathway to precipitation and precipitation patterns vary strongly depending on how shallow cumuli organise [211, 231]. Clouds which cluster tend to produce larger amounts of precipitation [294], with higher organisation sometimes increasing precipitation intensity by an order of magnitude [114]. Shallow cloud organisation is also the preamble for deeper convection with stronger precipitation. Therefore, regardless of whether or not (changing) organisation matters for Earth’s radiation budget, it is crucial for the spatial and temporal patterns of tropical precipitation. Moreover shallow cloud organisation is not only influential for warm-cloud microphysics, but also strongly dependent on it.

1.3.1.2 *The Role of Microphysics*

Observations seem to link precipitation from shallow cumulus to the formation of cold pools and their mesoscale organisation. Precipitating tropical shallow cumuli are ubiquitous and significant, with about one third of precipitation rates exceeding 1 mm h^{-1} [193], and rain is expected to evaporate efficiently because large-scale subsidence maintains a relatively dry boundary layer. Other observations show cold pools are also abundant [298], nearly always associated with previous precipitation [312], and are often associated with different mesoscale cloud patterns [298]. Like precipitation and cold pools, such patterns are also a trademark feature of shallow cumuli [e.g. 261]. Furthermore different cold pool characteristics, such as their frequency and cooling strength, covary with different patterns [298]. All this suggests rain evaporation-induced cold pools are in some way related to mesoscale shallow cumulus organisation.

Observational evidence is supported by LES which show rain evaporation is a primary driver of mesoscale cloud organisation. As rain falls through subsaturated air some of it evaporates. This evaporation moistens and cools locally, but more importantly triggers strong downdraughts which spread out as density currents when they reach the surface, forming cold pools [e.g. 120, 199, 236, 250]. All in all, rain evaporation induced cold pools stabilise the atmosphere and de-homogenise surface temperature and moisture fields, with cooling and drying in their centres suppressing clouds, “cloud-holes”, contrasted by moisture accumulation at their edges creating “cloud-arcs”. How and if cloud-holes and cloud-arcs then transition to different mesoscale cloud patterns like those observed during EUREC⁴A is poorly understood. It certainly depends upon large-scale environmental conditions such as near-surface wind speed and lower-tropospheric stability [33, 231], as well as shallow mesoscale circulations,

possibly related to the spontaneous self-aggregation of moisture which is intrinsic to shallow cumulus fields before they precipitate [115]. Nevertheless, LES identify cold pools triggered by rain evaporation as the dominant feedback controlling mesoscale shallow cloud patterns [e.g. 7, 235, 299]: when clouds precipitate rain evaporates, this affects how the next generation of clouds organise and this in turn affects how they will precipitate. As a result, precipitation is not just controlled by shallow cumulus organisation, but also plays an active role in shaping it.

Due to the role of precipitation, LES of shallow cumulus organisation shows high sensitivity to warm-cloud microphysics. Changing the microphysics scheme can directly affect the strength of rain evaporation and consequently cold pool characteristics and mesoscale cloud organisation [5–7, 128, 153, 213, 235]. Furthermore, many parallels can be drawn between cold pools from shallow cumuli and convective outflows from tropical deep convection [235, 312], and so known sensitivity to microphysics in deep convection is probably relatable to shallow cumuli² [178, 180, 203]. A particularly prominent case surrounds the sensitivity of both shallow and deep convection to the uncertainty regarding collisional breakup, where LES indicate that by changing DSDs, breakup might affect rain evaporation appreciably enough to alter cold pool strength [178, 180, 203, 235, 236, 265]. Aside from changes to rain evaporation, alterations to microphysics can also impact cloud radiative effects, which has repercussions for the timing of different cloud patterns with respect to the diurnal cycle and thus subsequent effects on the degree of organisation [183].

1.3.2 A SDM for Tropical Shallow Cumulus Organisation

The importance of shallow cloud organisation and its high sensitivity to warm-cloud microphysics motivates using SDM to model it. Many of the knowledge and inherent model uncertainties discussed in Section 1.1.2 directly relate to rain evaporation and/or precipitation timing and therefore could translate into uncertainties in cloud organisation. As explained in Section 1.2.1, SDM eliminates many of these inherent model uncertainties and can be used to more easily explore knowledge gaps. Furthermore we saw in Section 1.2.2 evidence that this is already being benefitted from to resolve uncertainties in tropical shallow clouds. Thus it seems likely that SDM could eliminate or diminish many of the sources of uncertainty in microphysics which hinders the development of understanding of shallow cumulus organisation.

² You could also flip this argument and see understanding the role of microphysics in shallow cumulus organisation as a simple test-bed for theories of deep tropical convection [235].

At the same time, substantial biases in shallow cumulus organisation and precipitation exist in mesoscale-domain [LES](#) as well as [GSRMs](#). At the start of this chapter, we saw many examples in which the poor representation of precipitation in both [GSRMs](#) and [LES](#) can be attributed to the problematic features of conventional models for warm-cloud microphysics, and that large changes in precipitation can be induced by changing microphysics schemes in small domain [LES](#) and [GSRMs](#) [e. g. [17](#), [93](#), [140](#), [185](#), [220](#), [268](#), [272](#), [289](#)]. Given the interdependence of precipitation and shallow cumulus organisation, it is therefore possible that some of the errors in not just precipitation but also organisation could be rectified by a different, also more adequate, description of warm-cloud microphysics. Here [SDM](#) seems like a promising candidate, not only to understand if and how uncertainties in microphysics impact shallow cumulus organisation, but also to reduce biases in precipitation and organisation in mesoscale-domain [LES](#) and [GSRMs](#) compared to observations.

In light of the above, we desire a [SDM](#) for warm-cloud microphysics capable of being used to study not just shallow cumulus precipitation but also shallow cumulus organisation. In the following chapter, I summarise why we therefore needed a novel computational implementation of [SDM](#) and describe the one we created, including its performance and warm-cloud microphysics. Having created this [SDM](#), I explain in Chapter [3](#) how we used it in two different ways to constrain uncertainties in microphysics which relate to shallow cumuli organisation and precipitation. First, we ran a [SDM](#) comparison to evaluate the robustness of precipitation modelled by [SDM](#) against its potential sources of inherent model uncertainties. Having found the amount of precipitation to be fairly insensitive to [SDM](#) model parameters, in the second study we combined the model with [EUREC⁴A](#) observations and better constrained the amount of, and controls on rain evaporation in the sub-cloud layer of shallow cumuli. We conclude in Chapter [4](#) reflecting on the progress the work in this dissertation has made and what possibilities it has opened up for the future.

*There, peeping among the cloud-wrack
above a dark tor high up in the mountains,
Sam saw a white star twinkle for a while.*

— J. R. R. Tolkien [278]

2.1 CREATING A SDM FOR LARGE DOMAIN SIMULATIONS

The computational cost of existing SDMs limits them to domains too small to resolve shallow cumulus organisation. As discussed in Section 1.3.2, we seek to use SDM as the microphysics scheme in mesoscale domain LES of tropical shallow cumulus organisation. However their current costs in terms of both memory and time have prevented them exceeding domains $\mathcal{O}(10\text{ km})$, far smaller than the 20 – 200 km domains required to capture mesoscale organisation [165, 228]. A pioneering SDM simulation was run in a mesoscale domain ($144\text{ km} \times 144\text{ km} \times 4\text{ km}$), but this was at such large computational expense that it limits the repeatability of such experiments and is already beyond the practical limit for most research (T. Yanase, personal communication, Nov 4, 2025). To use SDM for investigating the role of microphysics in tropical shallow cumulus organisation, we need it to be more computationally feasible in such large simulations. Hence we created a new SDM which is specifically suited for modelling warm-clouds on exascale (i. e. extremely high-performance) computers. Our novel SDM implementation is called Cleo, named as a tribute to the anonymous mathematician as well as Cleopatra VII Thea Philopator. This chapter is a summary of the two model description papers, Bayley et al. [22] and Bayley et al. [23], which we currently have under review at *Geoscientific Model Development* and are found in Appendix A and Appendix B.

To make Cleo fit for purpose for studying warm-clouds in domains large enough to resolve mesoscale shallow cumulus organisation, we created Cleo with three characteristic features: its memory layout, resource allocation, and flexibility. From previous work optimising SDM we identified that the largest barriers to a high-performance were the memory footprint of SDM and the cost of compute and of particle transport (Sec. 1.2.3) [11, 19, 57, 58, 165]. We therefore designed a memory layout which optimises memory allocation and access patterns, and we designed Cleo’s resource allocation to be efficient, meaning Cleo could harness compute from both GPUs and Central Processing Units (CPUs) across a range of different computer

architectures, and with the freedom to adjust load balancing. As well as these computational characteristics we wanted Cleo to facilitate microphysical process understanding and so we built Cleo with an extraordinary degree of flexibility. We achieved this primarily by taking the principles of monoids from mathematics and creating a computational analogy. Our monoids enable adaptive time-stepping, avoid conditional code branching, and make microphysics and data output highly configurable. Additionally, we built flexibility into the model where we saw the major knowledge gaps as discussed in Section 1.1.2.2, for example making it possible to change the effect of turbulence, and collisional breakup. The following section describes all these features of Cleo in more detail. Section 2.3 thereafter summarises their evaluation in terms of Cleo’s computational and microphysics performance.

2.2 MODEL DESIGN

2.2.1 Memory Layout

2.2.1.1 Superdroplets and Grid-Boxes

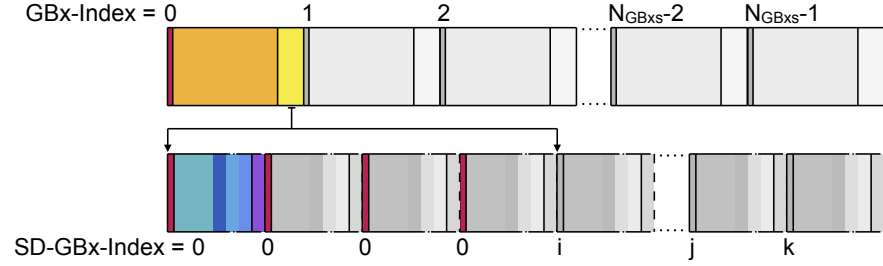


Figure 2.1: Schematic of the Array of Structures (AoS) memory layout of grid-boxes and superdroplets. The colours indicate the sub-components of the structure defining a single grid-box and superdroplet. For illustration here the first four superdroplets are associated with the first grid-box, indicated by GBx-Index = 0 matching their SD-GBx-Indexes (also coloured magenta). *This figure is reproduced from Bayley et al. [22] (App. A Sec. A.3).*

The memory layout of both grid-boxes and superdroplets in Cleo targets both efficient memory allocation and access patterns. The entire simulated domain is divided into sub-domains distributed across MPI processes (Message Passing Interface (MPI); see Sec. 2.2.2.2) and each sub-domain is composed of grid-boxes, which contain for example thermodynamic variables, and, separately, superdroplets. We chose to arrange both superdroplets and grid-boxes in an AoS layout, i.e. the grid-box/superdroplet structures are stored sequentially in a single array (Fig. 2.1). By choosing to store both superdroplets and grid-boxes in single contiguous memory allocations we reduce the

overhead from buffer memory compared to scattered data, and by storing superdroplets separately to grid-boxes we minimise the frequency and size of memory de-/re-allocations when superdroplets move around the domain. Additionally the superdroplets' array is ordered by the grid-box each superdroplet is associated with, such that superdroplets associated with the same grid-box are stored in a contiguous chunk of memory. This ordering makes access patterns efficient when performing loops over superdroplets, for example when enacting microphysical processes and superdroplet motion, because both grid-boxes and superdroplets are accessed sequentially, thereby minimising cache misses¹. Cache misses are also avoided during the [SDM](#) collision algorithm because with this layout superdroplets which are allowed to interact are close to each other in memory.

Although Cleo's memory allocations and access patterns have several advantages, using an ordered [AoS](#) memory layout for superdroplets comes with a price. During superdroplet motion superdroplet arrays become unordered and to rectify this a costly sorting algorithm must be applied. Furthermore, operations that act on a single superdroplet attribute can be inefficient because, in contrast with a Structure of Arrays ([SoA](#)) layout, Single Instruction, Multiple Data parallel processing is difficult to apply. Also cache misses for any calculation increase as the number of superdroplet attributes not involved in the calculation increases. This makes complicated superdroplets with many attributes less desirable. More comprehensive benchmarking is needed to confirm whether the benefits of the [AoS](#) layout outweigh the costs compared to the [SoA](#) layout and in reality some degree of compromise between the two (formed from an Array of Structures of Arrays) would be the better than either extreme. Nevertheless, the choices we made for Cleo's memory layout are consistent with increasing attention in [HPC](#) given to memory bandwidth optimisations and they are sensible for simulations involving many superdroplets with few attributes, exactly as we require for simulations of mesoscale shallow cumulus organisation.

2.2.1.2 Data Output

As well as during simulations, Cleo is designed to save memory in its data output. Data written to non-volatile memory is stored according to the Zarr storage specification version 2, which is well-suited to large N-dimensional arrays and parallel reading/writing. Additionally all data is cast to 4-byte types by default, approximately halving the memory footprint, and superdroplets' data are stored according

¹ A cache miss is when an instruction requires data that is not currently stored in the local, and hence small but fast, cache memory, and so must be retrieved from a larger, slower cache or random access memory which is yet larger, yet slower. Such retrievals take time and therefore cause latency, i.e. delays, whilst the instruction waits for the required data. Cache misses are therefore especially costly on large high-performance computers.

to the CF Metadata Convention’s contiguous ragged array representation. This ragged array format is particularly efficient for storing arrays whose lengths vary over time, exactly as occurs when superdroplets leave or enter the domain. Unlike regular arrays which need to keep the length of all dimensions fixed, a ragged array does not need to store superfluous data when superdroplets are not present.

2.2.2 Resource Allocation

To exploit the available compute on exascale computers, we designed Cleo not only to allocate memory, but also computational resources efficiently. This resource efficiency comes from Cleo’s coupling to a host dynamical driver, which facilitates load balancing, and its use of Kokkos and [MPI](#) libraries to harness the different types of shared and distributed memory parallelism available across different high-performance computers.

2.2.2.1 Coupling to a Host Dynamical Driver

The way that Cleo couples to a host dynamical driver enables independence of their grid and domain decomposition which can be used to optimise resource allocation. Cleo runs in tandem with a host dynamical driver which is responsible for determining the thermodynamics and wind fields that Cleo uses to evolve superdroplets; for example the driver may provide Cleo’s input from a file, or evolve it using a fluid-dynamical model such as ICOSahedral Nonhydrostatic (Model) ([ICON](#)). For maximum possible freedom, Cleo has the option to couple to a driver through the coupler called Yet Another Coupler ([YAC](#)). [YAC](#) enables the choice of grid and domain decomposition of the driver to be completely independent of Cleo. Therefore both Cleo and the driver can choose, without compromise, the grid and decomposition which optimises their own performance. [YAC](#) also allows for Cleo to run concurrently to the driver, thereby allowing another way to potentially save run-time. Most importantly however, the independence enabled by the coupling facilitates load balancing and thus efficient allocation of computational resources.

I implemented this coupling and Cleo’s distributed memory parallelism together with research software engineers, Wilton Loch and Aparna Devulapalli, during two National Earth System Modeling (Project) ([NatESM](#)) sprint projects which I applied for during my PhD.

2.2.2.2 Distributed and Shared Memory Parallelism

Cleo utilises distributed memory parallelism through [MPI](#) domain decomposition. The decomposition it chooses distributes sub-domains of the entire simulated domain so that grid-boxes are as evenly distributed as possible, but that vertical columns of the domain are pref-

entially kept on a single MPI process. Communication via MPI then only occurs during superdroplet transport, to transfer superdroplets which leave/enter a sub-domain. Since the most frequent motion of superdroplets is usually vertically, by choosing the domain decomposition is this way we avoid excessive communication. Nevertheless, we have not yet evaluated the performance of Cleo’s MPI domain decomposition because the most suitable decomposition for high-performance almost certainly depends on the fluid-flow given by the host dynamical driver as well as the specific computer architecture. This undoubtedly leaves room to improve Cleo’s particle transport, as will be discussed in Chapter 4.

To exploit shared memory parallelism, also across different high-performance computers, Cleo uses the Kokkos libraries. Kokkos enables us to write a single source code for Cleo which builds for a diverse range of computer architectures, including both CPUs and GPUs [60]. At compile-time, Kokkos evaluates the available hardware and any optional user-defined settings to determine the most suitable way to allocate resources to achieve high-performance. For example, Kokkos determines task sizes for parallel loops appropriate for the cache of the given architecture or execution resource (CPU or GPU). Combined with the superdroplets’ memory layout explained in Section 2.2.1, this amounts to optimisation by cache blocking. By using Kokkos we therefore implement shared-memory parallelism with high-performance portability. The results of this parallelism are presented in Section 2.3.2.

2.2.3 Flexibility

As stated earlier, flexibility is Cleo’s third characteristic feature, specifically to facilitate microphysical process understanding. We created such flexibility by exploiting C++ templating, both to represent knowledge gaps and to construct microphysical processes and data output which behave like monoids.

2.2.3.1 Through Monoids

In mathematics, a monoid is a closed set with an associative binary operation and an identity element. Put simply, when two elements (members) of a monoid set are added together their outcome is another element of the set, and the identity element is defined such that the outcome of its addition is just the other element involved in the addition unchanged.

We applied the idea of monoids computationally to construct “observers” for data output and microphysical processes. Taking microphysics as the example, we defined the monoid set for microphysical processes by some constraints a code structure must satisfy to be allowed to be a member of it. To define the associative binary op-

eration, we created rules for how two such code structures should be combined to produce a new microphysical process which is the “sum” of its parts (associativity, but not commutativity). Finally for the identity element, we designed a code structure which both obeys the constraints for the microphysical process set and does not change anything when it is combined with another element of the set according to our rules for the associative binary operation.

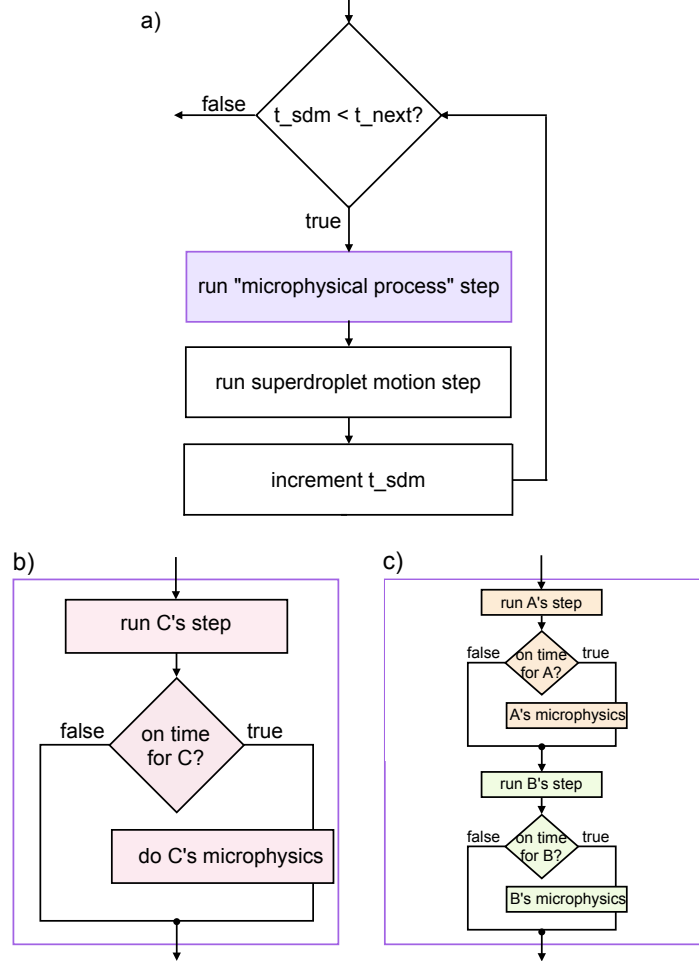


Figure 2.2: Schematic illustrating how the microphysics enacted at run-time can be composed of any number or permutation of different individual microphysical processes. (a) During the [SDM](#) sub-time-stepping routine at run-time, there is a call to run a “microphysical process” whose content depends on what it was instantiated with during compilation; (b) An example of the content if it was instantiated with a single microphysical process “C”; (c) An example if it was instead instantiated with the sum of two individual processes, “A” and “B”. Extending the schematic, “A” and “B” could be combined in a different order, and/or with other microphysical processes. (For the full Cleo time-stepping routine, see App. [A](#) Fig. [A.9](#)).

To implement all this computationally we chose to use C++ templating in combination with the recently introduced C++ definition

of Concepts [112]. Specifically, we created a C++ Concept which enforces the set of constraints for a microphysical process and then we created a function template which implements the rules for how to combine any C++ Structures/Classes which obey these constraints. In this way, we can construct various different microphysical processes and then, during compilation, any number and permutation of them can be combined to instantiate a templated type. This instantiation, itself a microphysical process, is used at run-time to enact microphysics (Fig. 2.2). The same is true for observers and could be easily extended to enact other processes in the future, aside from microphysics or data output; some possibilities are discussed in Section 4.2.1.

Several computational benefits come from implementing our monoids through templates and C++ Concepts. By choosing microphysics at compile-time we avoid wasting run-time first by not compiling unused code and hence increasing code locality and second by removing chances for incorrect branch prediction at points of conditional code-branching. Additionally compilers can efficiently inline and optimise templated code and give more comprehensible error messages than without C++ Concepts.

More importantly, our monoids make microphysics and data output highly configurable, aid code maintainability, and enable adaptive time-stepping. The principal advantage of monoids is the extraordinary degree of flexibility enabled for microphysical process studies. Users can easily design experiments which enable different versions of microphysics and they can tailor data output specifically to their needs too. Whilst this could in principle be possible with conditional code-branching, it would quickly produce an unwieldy, unsustainable code-base which would be error prone and hard to maintain. Furthermore, we designed our microphysical process and observer monoids such that they enable adaptive time-stepping, which means the combined microphysical process/observer respects the time-step of each individual element it is composed of whilst also always taking the largest possible time-step. Hence the flexibility in Cleo which we created via monoids promotes code maintainability and run-time performance as well as being purpose-fit for microphysical process studies.

2.2.3.2 To Represent Knowledge Gaps

Cleo has built-in flexibility to account for major knowledge gaps regarding turbulence-microphysics interactions and collisional breakup. As discussed in Section 1.1.2.2, a major source of knowledge uncertainty is turbulence-microphysics interactions. To reflect this uncertainty in Cleo, we use a templated type to represent the collision kernel which is used to calculate the probability two superdroplets collide. This means collision kernels can be easily interchanged, for example to neglect the effect of subgrid-scale turbulence (as is con-

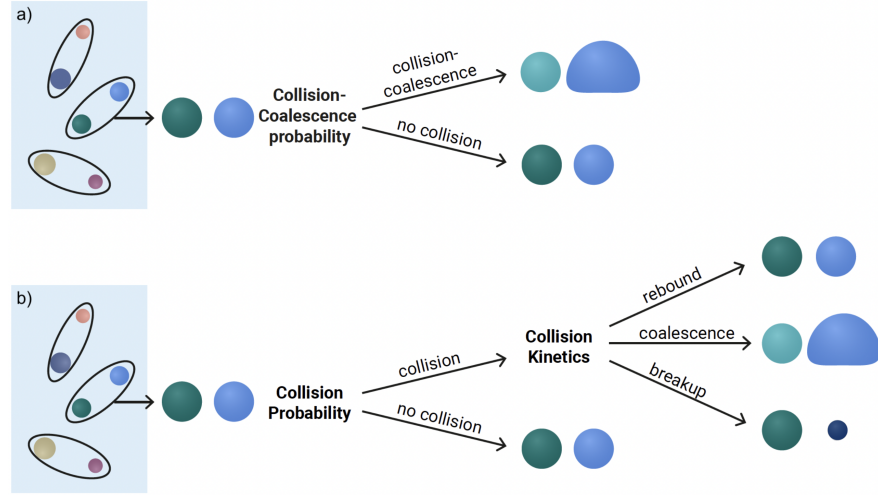


Figure 2.3: Schematic of the two treatments for collisions available in Cleo. (a) The original collision-coalescence algorithm from Shima et al. [241]; (b) the extended collision algorithm including breakup and rebound as well as coalescence. A lightening/darkening of a superdroplet indicates a decrease/increase in its multiplicity. This figure is reproduced from Bayley et al. [23] (App. B Sec. B.4).

ventionally done) or to account for its effect according to one of the various ways described in the literature. Likewise we template over the calculation of superdroplet motion so that it too could easily be extended to include the possible effects of turbulence. To account for the large uncertainties but potential importance of collisional breakup (Sec. 1.1.2.2 and Sec. 1.3.1.2), we designed two alternative treatments for Cleo’s collision algorithm (Fig. 2.3). The first treatment is the original collision-coalescence algorithm proposed by Shima et al. [241], the second is an extended version which can further represent rebound and breakup. In relation to the discussion of monoids above, these treatments are implemented as two different microphysical processes and so can be easily interchanged during compilation of the model. Our extended algorithm allows us to explore the range of uncertainty over the potential effect of collisional breakup and it has notable differences from the other SDM collisional breakup algorithms. Unlike Bringi et al. [38], we ensure conservation of superdroplet number and thus computational tractability, and unlike De Jong et al. [51] we avoid an additional probabilistic step and a recursive algorithm. These differences reduce the computational cost of Cleo’s breakup algorithm compared to the existing alternatives, but still make it useful for assessing potential impacts of breakup, as will be demonstrated in the case of rain evaporation in Section 3.2.

2.3 MODEL EVALUATION

2.3.1 Microphysics Verification

Efforts in Cleo’s fundamental design are only worthwhile if the microphysics works as desired. To model all the major microphysical phenomena driving warm-clouds, we chose to implement standard algorithms from the literature, with some important adjustments. We took ideas from Shima et al. [241] and Matsushima et al. [165] to use an implicit Euler method with the Newton-Raphson root-finding algorithm to solve the ordinary differential equation for condensation/evaporation. This method models (de-/re-)activation and droplet growth/decline explicitly according to Köhler theory. Additionally however, we included a term into the equation to account for the ventilation effect, the importance of which will be shown in Section 3.2.

How Cleo can represent droplet collisions has already been explained above, and we have implemented various different collision kernels from the literature based on Golovin [75] (unphysical but useful for model validation), Long [159] (hydrodynamic for collision-coalescence only), Low and List [161], Szakáll and Urbich [270], and Schlottke et al. [230], Straub et al. [266] and Testik et al. [274] (different hydrodynamic formulations for collisions including breakup).

Finally, to simulate droplet motion, we followed the arguments of Grabowski et al. [78] and used Heun’s method, a second-order predictor-corrector method, with a simple linear interpolation of the wind velocity; however, as mentioned in Section 2.2.3.2, we template over this method to allow for future extensions of the algorithm which could account for subgrid-scale turbulence. In light of the slight discrepancies between different formulations for droplet terminal velocities, we also template over them which allows users to account for droplet sedimentation according to either Rogers R.R. [218], Rogers et al. [217], or Simmel and Wurzler [246].

We verified Cleo’s warm-cloud microphysics with four known test-cases. The first three test-cases each illustrate the behaviour of one of the warm-cloud processes Cleo can currently represent, namely condensation/evaporation, droplet collisions and droplet motion. The final test-case is an integrated test of Cleo’s capabilities using the 1-D KiD framework [243]. All four test-cases are presented in detail in Appendix B Section B.6, but here we summarise the two most notable.

We validated Cleo’s collision algorithms using the same test-case described in Section 5.1.4 of Shima et al. [241] (Fig. 2.4). We reproduce their results for evolution of the mass density distribution for collision-coalescence using the Golovin and hydrodynamic collision kernels. As well as that, we exemplified how the evolution differs when the hydrodynamic kernel is used with Cleo’s alternative collision

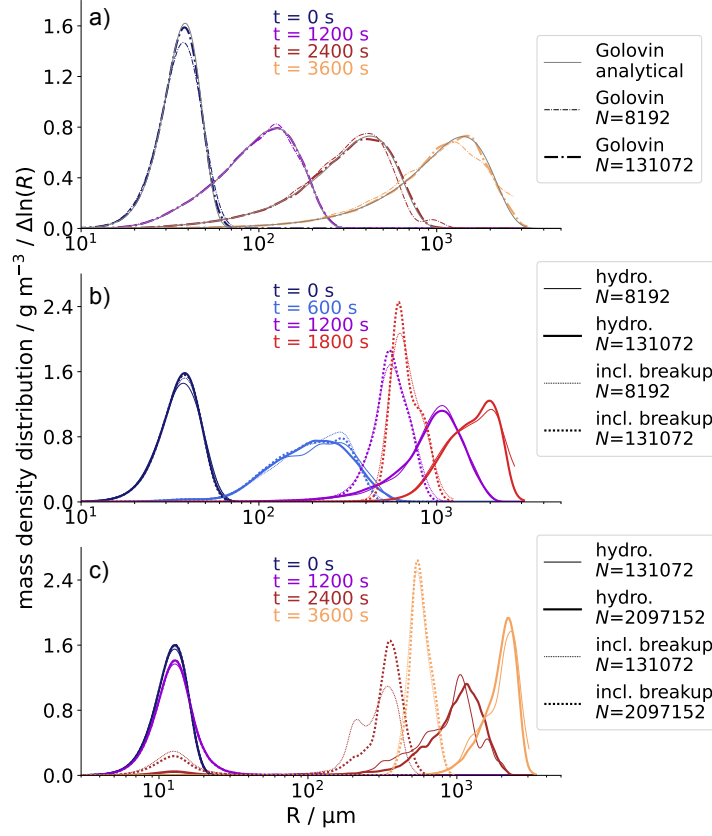


Figure 2.4: Test of **SDM** collisions as in Shima et al. [241, Sec. 5.1.4, Fig. 2]. The line-style indicates the collision kernel; the line-width indicates the number of superdroplets, N ; the colour indicates the time. Sub-figure (a) additionally shows the analytical solution for the Golovin kernel at each time (solid grey). The sub-figures are the same as in Shima et al. [241] except in panels (b) and (c), where, as well as the hydrodynamic kernel, we additionally plot the distribution's evolution when breakup and rebound are accounted for based on the combination of parametrisations from Schlottke et al. [230], Straub et al. [266], and Testik et al. [274]. This figure is reproduced from Bayley et al. [23] (App. B Sec. B.6).

treatment, including collisional breakup and rebound as well as coalescence. The parametrisation chosen here a combination of the works of Schlottke et al. [230], Straub et al. [266], and Testik et al. [274] (see App. B Sec. B.4.2). As expected, accounting for breakup delays the growth of droplets compared to the collision-coalescence-only case and mm-size droplets are inhibited. Unfortunately a more quantitative validation of the effect breakup on this test-case is limited by the large uncertainties surrounding its parametrisation.

The results from our integrated test lie within the spread amongst different **SDMs** in [94] (Fig. 2.5). As expected, for the setup only including condensation/evaporation and droplet motion without sedimentation, the liquid water content reaches steady state in the middle of the cloud without ensemble spread (Fig. 2.5a,b). When collision-

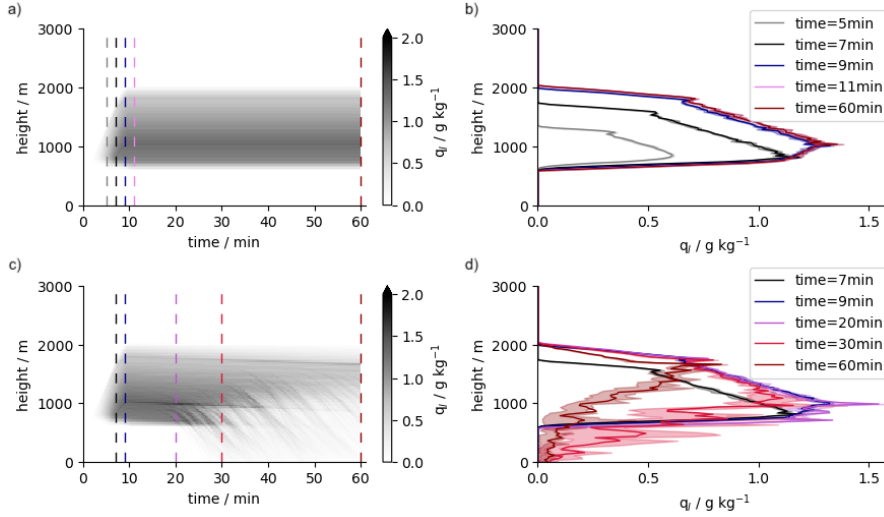


Figure 2.5: Evolution of the liquid water mass mixing ratio, q_l , during the 1-D Kinematic Driver (KiD) test-case: (a,b) only modelling condensation/evaporation, (c,d) modelling condensation/evaporation, collision-coalescence and droplet sedimentation. Results are shown for the mean and interquartile range over an ensemble of 10 members, and (b) and (d) are cross sections across the vertical coloured lines in (a) and (c) respectively. *This figure is reproduced from Bayley et al. [23] (App. B Sec. B.6).*

coalescence and sedimentation are also included, the stochasticity of the SDM collision algorithm introduces ensemble spread, and precipitation formation begins around 30 min (Fig. 2.5c,d), similarly to SDMs in Hill et al. [94]. From all the microphysics tests we conducted we can conclude that Cleo is a fully-functioning SDM capable of representing all the major microphysics of warm-clouds.

2.3.1.1 The Microphysics Test-Cases Repository

Alongside verifying Cleo’s microphysics behaviour, during my PhD I also collaborated with a group at Deutsche KlimaRechenZentrum (DKRZ) to create a generic test-case framework for evaluating any cloud microphysics scheme in known test-cases [27]. I took the test-case for Cleo’s condensation/evaporation and the 1-D KiD and rewrote them in Python with a standard interface such that any arbitrary microphysics scheme, given its own Python wrapper, could use them too. Using this framework, we can run these two test-cases for both Cleo and ICON’s refactored one-moment bulk scheme during continuous integration (e. g. on GitHub or GitLab). This can of course be extended to other microphysics schemes, and facilitates testing for the future development of both schemes in a semi-automated and consistent way.

2.3.2 Computational Performance

2.3.2.1 Strong- and Superdroplet- Scaling

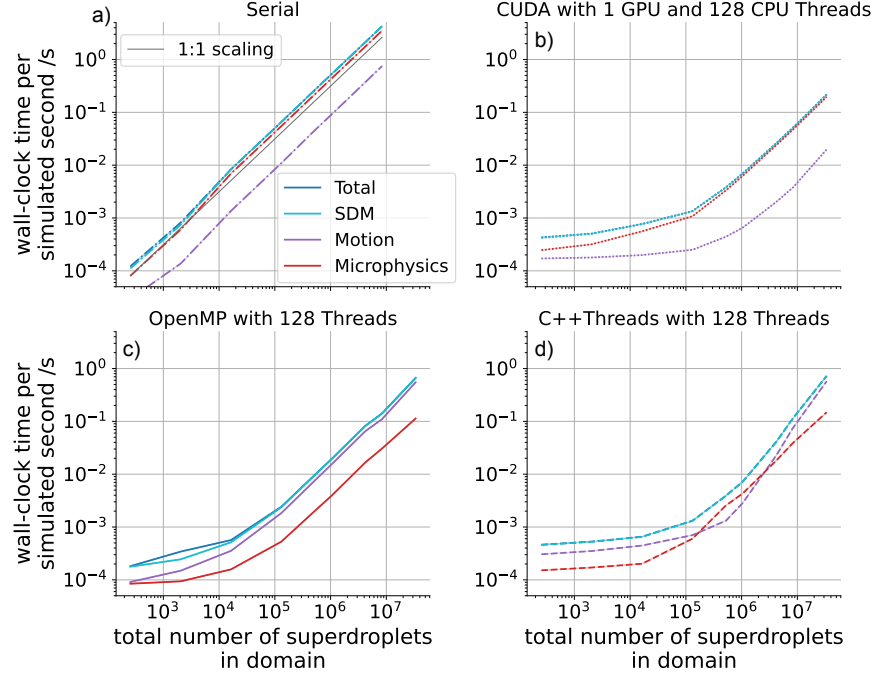


Figure 2.6: The superdroplet-scaling of the wall-clock time per simulated second given the maximum set of resources on Levante for each build: (a) Serial (dash-dotted), (b) CUDA (dotted), (c) OpenMP (solid), and (d) C++Threads (dashed). The colours decompose the total time (blue) into the time spent on **SDM** (cyan), motion (purple) and microphysics (red). *This figure is reproduced from Bayley et al. [22] (App. A Sec. A.7).*

Having verified the microphysical behaviour of the model, we conducted an evaluation of its computational performance on different shared memory architectures. The key measures of Cleo’s shared-memory performance are presented in Figure 2.6 and Figure 2.7. Cleo’s performance was tested on a single **CPU** or **GPU** node of Levante **HPC** system with simulations of a 3-D Cartesian $6 \text{ km} \times 0.3 \text{ km} \times 1.5 \text{ km}$ domain representing the major processes in warm-clouds: condensation/evaporation, collision-coalescence, and droplet motion including sedimentation. The details of the computer and simulation setup are described in Appendix A Section A.7. As desired, linear superdroplet-scaling is achieved on all forms of parallelisation and almost all the wall-clock time is spent in time-stepping **SDM** (Fig. 2.6). Cleo’s memory footprint also shows good scaling and a reasonable effective size of both a grid-box and superdroplet $\mathcal{O}(100 \text{ bytes})$, similar to their source-code values (App. A Fig. A.8).

The strong-scaling of the simulations does not reach optimal efficiency due to bottlenecks in the microphysics and motion algorithms

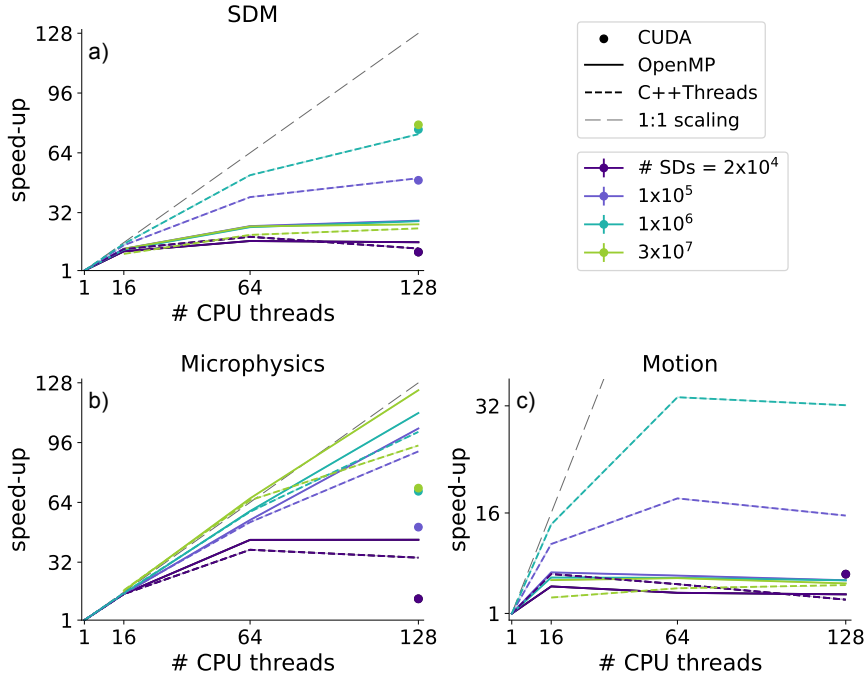


Figure 2.7: The strong-scaling of the speed-up of wall-clock time relative to the serial simulations with increasing number of CPU threads on Levante for (a) time spent on SDM, (b) microphysics, and (c) motion. The line-style denotes the build and the colours from purple to green indicate increasing problem size in terms of the total number of superdroplets in the domain. The speed-up of motion from CUDA exceeds the y-axis scale and is reported in Appendix A, Table A.2 instead. *This figure is reproduced from Bayley et al. [22] (App. A Sec. A.7).*

(Fig. 2.7). For CUDA parallelism, the bottleneck is the shuffling algorithm used during collision-coalescence, whereas for OpenMP and C++Threads it is the sorting algorithm required during superdroplet motion. Both bottlenecks are expected to become less significant with more complex simulations, particularly for large-domain simulations which will require frequent MPI communication. Hence it only makes sense to further optimise the shuffling and sorting algorithms, for example following the work of Decyk and Singh [52] for shuffling (as has already been proved successful [165]), after we have evaluated its cost in more complex simulations.

2.3.2.2 Parallels with Matsushima et al. [165]

Cleo displays many similarities to the most performant SDM to-date. Matsushima et al. [165] optimised the SDM developed by Shima et al. [241] specifically for Fugaku supercomputer and achieved exceptionally high-performance across three standard test-cases. Albeit differently, they also employed cache blocking, superdroplet sorting, and reduced code-branching to optimise their SDM, and superdroplet

transport is their bottleneck in agreement with our CPU-only performance. These similarities bode well for the performance of Cleo in simulations with MPI domain decomposition, and also point towards possible areas for further optimisation, for example using single precision for some superdroplet attributes during run-time. We also note that Cleo has major additional benefits from its performance portability and model flexibility.

2.4 SUMMARY

This chapter introduced a novel computational implementation of SDM called Cleo. Cleo was created to address demand for a SDM fit for the purpose of studying warm-cloud microphysics in domains large enough to resolve mesoscale shallow cumulus organisation. We therefore focussed our development on three main characteristics so that Cleo can be computationally efficient on exascale computers, and at the same time facilitate microphysical process studies. These three characteristics are Cleo's memory layout, resource allocation, and its flexibility. As a result, Cleo shows optimal superdroplet-scaling and sufficient strong-scaling on a single node to be ready for larger simulations and the optimisation of its particle transport. On top of this, Cleo is capable of representing all the principal microphysical processes of warm-clouds, including collisional breakup, and it accounts for knowledge uncertainties in those representations. Having created a viable SDM for warm-cloud microphysics, I explain in the following chapter how we have begun using it for better constraining uncertainties relevant to shallow cumulus organisation. In particular we tightened constraints on how much inherent model uncertainty from SDMs impacts the behaviour of precipitation from warm shallow clouds, and how much rain evaporation occurs in sub-cloud layer of shallow cumuli and what controls it.

USING CLEO TO CONSTRAIN UNCERTAINTIES IN WARM-CLOUD MICROPHYSICS

*At some point we have to give up and say
that's just the way it is.
Or, not give up and push on.*

— L. Susskind [269]

3.1 CONSTRAINING INHERENT MODEL UNCERTAINTY IN SDM

3.1.1 *A Model Inter-Comparison to Constrain Inherent SDM Uncertainties*

It is unclear to what extent inherent model uncertainties in [SDM](#) impact its representation of warm-clouds. As discussed in Section [1.1.2.3](#), conventional bulk/bin models for cloud microphysics suffer from numerous inherent model uncertainties and these contribute to large uncertainty in both [LES](#) and [GSRMs](#) for the macrophysical outcomes of warm-clouds (Sec. [1.1.1](#)). One of the main motivations for using [SDM](#) to model shallow clouds therefore, is to eradicate this uncertainty arising from bulk and bin models (Sec. [1.2](#)). However, because [SDM](#) still stipulates a few somewhat arbitrary modelling choices, albeit different ones, using [SDM](#) does not a priori mean the reduction of uncertainty stemming from the way we model cloud microphysics. [SDM](#) is only useful to us and for reducing inherent model uncertainties if the macrophysical properties of warm-clouds are less sensitive to the model-related choices of [SDM](#) than to those of conventional bulk/bin models.

Previous studies motivate further investigation of the sensitivity of warm-clouds to certain modelling choices in [SDM](#). Several studies have pointed out [SDMs](#) can only model precipitation reliably if the number of superdroplets per grid-box exceeds $\mathcal{O}(10)$ to $\mathcal{O}(100)$ [[177](#), [286](#), [287](#)]. However, on top of this Hill et al. [[94](#)] ([H2023](#)) suggested that warm-cloud precipitation may be highly sensitive to two further sources of inherent model uncertainty in [SDM](#). They used the 1-D Kinematic Driver ([KiD](#)) framework [[93](#), [243](#)], to compare microphysics schemes under prescribed dynamics and found less consensus between two [SDMs](#) than bulk and bin models. This was across a range of metrics but notably included precipitation amount and onset. They suggested the differences could have been caused by “(i) the size initialization of the superdroplets, (ii) interpolation of collision efficiencies, and/or (iii) numerical convergence with the number of collision

pairs per second.”¹ and therefore concluded that we “do not have a satisfactory numerical benchmark for precipitation processes”. The advent of [SDM](#) had raised the expectation that [SDM](#) would alleviate inherent model uncertainty and therefore enact more consistent microphysics than conventional models, yet in that study the comparison of two [SDMs](#) seemed to directly contradict this notion. This is of particular concern to us because if modelling choices in [SDM](#) impact precipitation, they would confound efforts to use [SDM](#) to understand shallow cloud organisation (Sec. 1.3.2). Given the gravity of this, not only as a challenge to one of the most important expected advantages of [SDM](#), but also to our goal of using [SDM](#) to better constrain the role of precipitation and its evaporation in shallow cumulus organisation, we found it necessary to further assess these statements’ validity.

We used 1-D [KiD](#) models for two different [SDM](#) implementations, Cleo and PySDM [21, 123], coupled to the same dynamical driver, PyMPDATA [20], to explore the potential sources of uncertainty in precipitation from [SDMs](#), including the three sources suggested by [H2023](#). We choose to deviate slightly from the strict specification of the [KiD](#) model which specifies the implementation of the dynamical driver from Shipway and Hill [244], and we improved on the boundary conditions for superdroplets from [H2023](#) (App. D Sec. D.3.3). Notwithstanding these changes, our [SDMs](#) couple to the same dynamical driver equivalently to previous [KiD](#) studies; therefore our [KiD](#) models using Cleo and PySDM have the same advection scheme and prescribed-flow, prescribed-temperature thermodynamic setup. The [KiD](#) thermodynamic setup stipulates a 3 km column model with a constant potential temperature profile and an initial relative humidity profile and concentration of [CCN](#). A prescribed updraught velocity for the first 10 min then advects water vapour, [CCN](#) and liquid droplets, but not temperature or dry air, which causes droplet activation and growth in the upper levels of the column followed by precipitation.

We ran ensembles of 10 members using Cleo and PySDM and analysed the sensitivity of the Liquid Water Path ([LWP](#)) and precipitation in the column to various different modelling choices. To investigate any sensitivity of precipitation to the size initialisation of superdroplets, we used the alpha-sampling method from Matsushima et al. [165], whereby a different value of the single parameter “alpha”, α , constructs a different choice of initialisation, as demonstrated in Figure 3.1 (App. D Sec. D.3.3). To test sensitivity to collision efficiencies, we used Cleo with a realistic parametrisation based on Long [159], and compared this to Cleo and PySDM without collision efficiencies (i. e. collision efficiencies equal to 1.0 for all droplets). Finally, to resolve any discrepancies arising from different degrees of numer-

¹ Size initialisation in this context means the method used to sample the initial aerosol distribution to initialise superdroplet radii. Collision efficiencies here means the efficiency with which two droplets collide and coalesce according to the hydrodynamic kernel, see e. g. Shima et al. [241] equation 4.

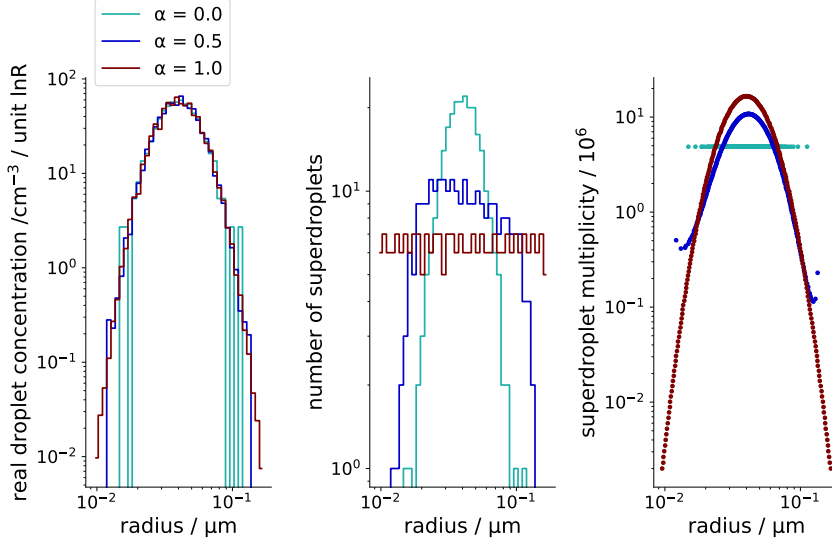


Figure 3.1: An example from this study of the different ways superdroplets can be initialised to represent an aerosol distribution via the alpha-sampling method. With $\alpha = 0.0$ superdroplets are initialised with equal multiplicities and their distribution of sizes are randomly sampled to recreate the shape of the aerosol distribution; with $\alpha = 1.0$ superdroplets are initialised with approximately equidistantly spaced sizes in log-radius space and their multiplicities sample the shape of the aerosol distribution instead. Intermediate values of α lie between these two end-points, as exemplified here by $\alpha = 0.5$. Thus by changing the value of α one can seamlessly move between different methods of superdroplet initialisation. Also visible here is that lower values of α less well sample the tails of the aerosol distribution. *This figure is reproduced from App. D Sec. D.3.3.*

ical convergence across [SDMs](#), we explored the effect of increasing the number of superdroplets per grid-box in the column. In these ways we sought to test the significance of the speculated sources of inherent model uncertainty and revisit whether or not [SDMs](#) can achieve consensus for the behaviour of precipitation from warm-clouds, and therefore be useful as a numerical benchmark for microphysics model development as well as for studying shallow cumulus organisation. A manuscript detailing the early results from our work, of which I am first author, is currently in preparation and can be found in [Appendix D](#).

3.1.2 Results of the 1-D KiD with Cleo and PySDM

We find that the sources of spread amongst [SDMs](#) cannot be attributed to differences in superdroplet initialisation and it is implausible that realistic collision efficiencies cause significant differences (Fig. 3.2). When we do not use collision efficiencies in both Cleo and PySDM, regardless of the value of α we observe that precipitation onset differs

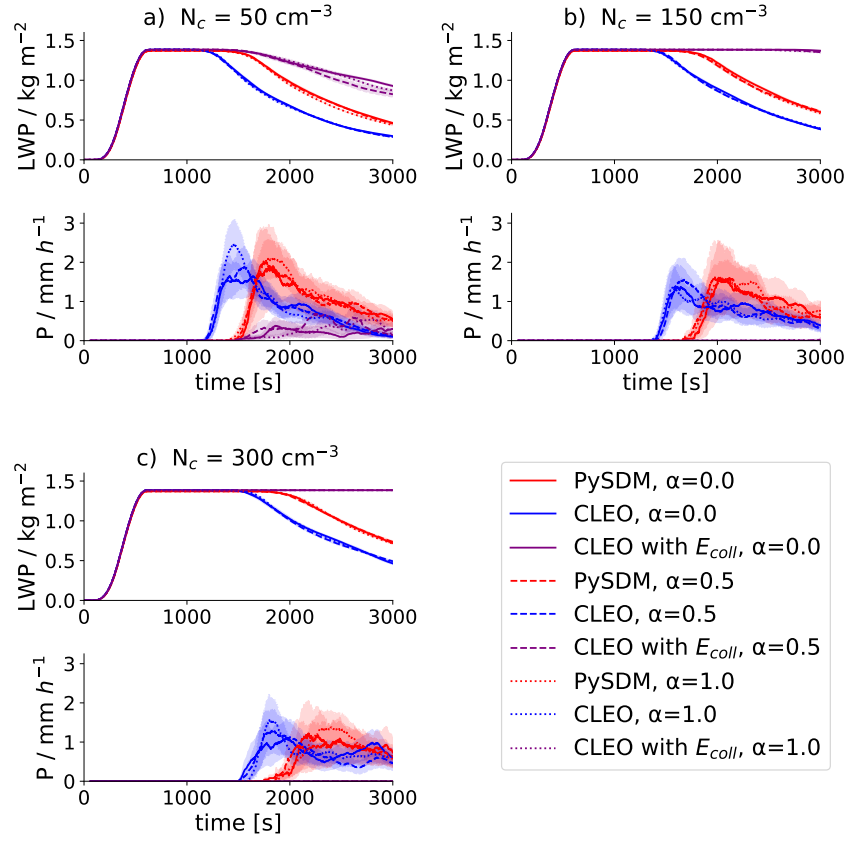


Figure 3.2: LWP and surface precipitation as a function of time for three different aerosol loadings, N_c , (a) 50 cm^{-3} , (b) 150 cm^{-3} , and (c) 300 cm^{-3} . The different lifestyles indicate different superdroplet initialisation methods via different values of α . The different colours indicate different microphysics: Cleo with parametrised collision efficiencies, E_{coll} (purple), Cleo without collision efficiencies (blue), and PySDM without collision efficiencies (red). This figure is reproduced from App. D Sec. D.4.

by almost 10 min. This shows that precipitation in our KiD models is sensitive to differences unrelated to the sampling method or collision efficiencies. As will be explained below and in Section 3.1.3, these other differences are also a more plausible explanation for the spread amongst SDMs seen in H2023. With regard to the amount of spread that can be caused by collision efficiencies, by comparing Cleo using realistic collision efficiencies to Cleo using none, we can see that under the cleanest aerosol conditions the maximum amount of precipitation decreases by an order of magnitude and precipitation onset is delayed by approximately 10 min (co-incidentally a similar amount as the difference between PySDM and Cleo without collision efficiencies). In scenarios with higher aerosol loading, not using collision efficiencies can even trigger precipitation which would not occur when using a realistic parametrisation. These results indicate precipitation modelled by SDM can be sensitive to the choice of parametrisation

of collision efficiencies. However, we hypothesise that the difference in precipitation as a result of using two different realistic parametrisations, rather than between using and not using a parametrisation, would be much smaller and therefore could not explain the large differences reported by H2023. Further work is on-going to test our hypothesis. With regard to changing the method of sampling used during superdroplet initialisation, changing the value of α makes almost no difference to the LWP and precipitation for both Cleo and PySDM; and any difference is certainly negligible within the ensemble spread. Our results therefore suggest precipitation is more robust to the choice of superdroplet initialisation and collision efficiencies in SDM than previously thought, which rules them out as major sources of inherent model uncertainty.

In addition, we found that SDMs are highly sensitive to ambiguities in the prescription of thermodynamics in the 1-D KiD setup. The setup described in Shipway and Hill [243] and H2023 defines the potential temperature and initial relative humidity profile, but this leaves subtle ambiguities about how to derive the pressure, density, and temperature profiles which SDM requires to enact microphysics. We therefore tested the sensitivity of precipitation to these ambiguities by running Cleo with parametrised collision efficiencies whilst making different assumptions for the thermodynamic derivations (Fig. 3.3). These were: not accounting for the effect of water vapour on the pressure profile, deriving the pressure profile from the Exner function instead of hydrostatic equilibrium, and additionally assuming constant density for the updraught velocity calculation. Superficially, these changes seem small and therefore could be easily overlooked, however they significantly affect the supersaturation of the cloud layer which has consequences for the DSD evolution modelled by SDM and hence LWP and precipitation (such large effects on the DSD may equally well be apparent in bin schemes although this is beyond the scope of our study). Because SDM shows high sensitivity to relatively small differences in the prescribed thermodynamics, it is plausible that similarly small differences, for example related to the calculation of supersaturation within different SDMs, could explain the large spread across SDMs in H2023 and the remaining difference between PySDM and Cleo without collision efficiencies. This is plausible even though the SDMs couple to the same dynamics driver (in our case PyMPDATA, in H2023 the driver from Shipway and Hill [244]). At the very least subtle differences in the thermodynamic setup appear to be a more plausible reason for large differences between SDMs in the KiD than superdroplet initialisation, collision efficiencies, or, as will be explained next, the degree of convergence of precipitation with respect to the number of superdroplets per grid-box.

Overall, our first results using the 1-D KiD framework indicate that two SDMs could agree more closely than previous work suggested.

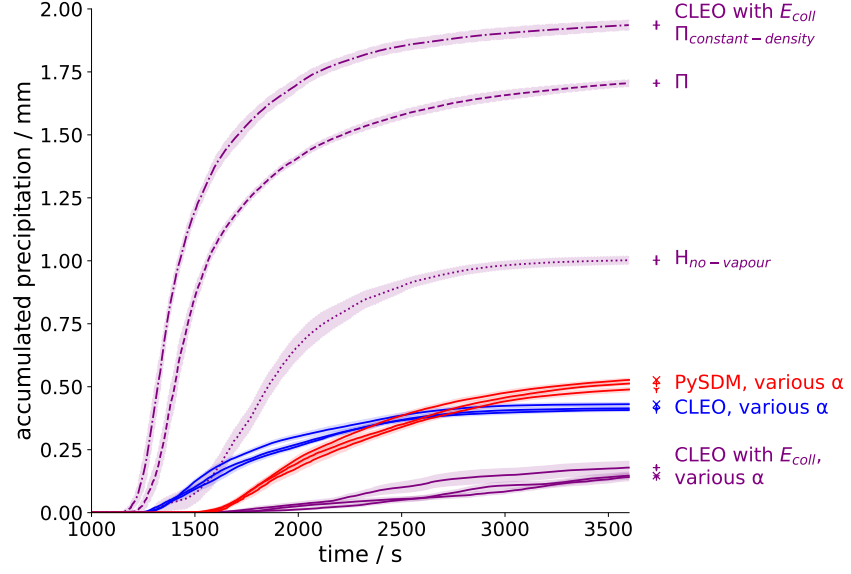


Figure 3.3: Accumulated precipitation as a function of time with 50 cm^{-3} initial aerosol loading under various different microphysical and thermodynamic setups. The solid lines refer to the standard hydrostatic setup as in Figure 3.2 for Cleo with parametrised collision efficiencies, E_{coll} , (purple), Cleo without collision efficiencies (blue), and PySDM without collision efficiencies (red); The different marker styles denote three different superdroplet initialisations: $\alpha = 0.0$ ($\dagger$), $\alpha = 0.5$ ($+$), and $\alpha = 1.0$ ($\times$). The other line-styles refer to the setups of Cleo with E_{coll} (and $\alpha = 0.5$) under perturbed thermodynamics: hydrostatic without accounting for water vapour, $H_{\text{no-vapour}}$ (dotted); thermodynamics instead calculated with the Exner function, Π (dashed); and thermodynamics calculated with the Exner function but also with a uniform density profile, $\Pi_{\text{constant-density}}$ (dash-dotted).

From Figure 3.2 and Figure 3.3 (red and blue lines) we can see that the precipitation rate and total amount of precipitation modelled by Cleo and PySDM very nearly agree with one another within ensemble error, except that, as mentioned earlier, the triggering of precipitation in Cleo is almost 10 min earlier than in PySDM. We are still working to understand the cause of this difference as discussed in the following section. Nevertheless, the pattern of precipitation is reassuringly extremely similar between the models and we are confident we will be able to understand and rectify the remaining discrepancy. Once we have, we will be able to show the consistent “solutions” from Cleo and PySDM for our **KiD** setup. Furthermore, by increasing the number of superdroplets per grid-box in Cleo we found the mean of precipitation is already converged at $\mathcal{O}(10)$ superdroplets per grid-box, which rules out superdroplet number as a source of spread in **H2023**. As well as this, because the variance decreases sufficiently quickly, at $\mathcal{O}(1000)$ we would be able to use **SDM** to provide a solution which

has a sufficiently small ensemble error to be useful as a numerical benchmark for comparing with other microphysical models.

3.1.3 *Summary*

Using Cleo and PySDM we have found that precipitation modelled by SDM in the KiD setup (prescribed-flow, prescribed-temperature) is much less affected by inherent model uncertainty than previous work suggested. Specifically, we have found it is robust to three arbitrary modelling choices in SDM: the number of superdroplets per grid-box, superdroplet initialisation, and the choice of collision efficiencies. Precipitation shows negligible sensitivity to the choice of superdroplet initialisation, and is already sufficiently converged at $\mathcal{O}(10)$ superdroplets per grid-box. Our early results testing precipitation's sensitivity to collision efficiencies leads us to conclude it is implausible they can explain the large differences between SDMs reported in H2023. Moreover, we have found other large sensitivities related to ambiguities in the thermodynamic setup which suggest subtle differences between thermodynamics could be a more likely explanation for the differences between SDMs reported in that study. We conclude that not only does SDM eradicate many of the sources of model uncertainty which conventional bulk/bin microphysics models suffer from, but also that the macrophysical properties of warm-clouds, in our case LWP and precipitation, can be robust to three potentially large sources of model difference in SDM. Our results therefore lend support to one of the key expectations of SDM, that it can reduce uncertainty in climate models by reducing inherent model uncertainty.

As well as reducing model uncertainty, our results support further work towards the use of SDM as a numerical benchmark for warm-cloud microphysics. We are able to provide a consistent solution for the pattern of precipitation across two different computational implementations of SDM except that the timing of precipitation is still not within error. Nevertheless, we are confident that with further work we can diagnose and rectify this discrepancy, and this in itself will bring further insight into what precipitation modelled by SDMs is sensitive to. We have already ruled out many potential factors and, given the sensitivity of the SDMs to the details of the KiD setup, we are working to understand if remaining differences in thermodynamics are playing a role. Besides this, we are exploring whether a parametrisation not discussed here, for example the relative fall velocities of the droplets, could be responsible for hastening the DSD evolution in Cleo compared to PySDM. Whether the difference is SDM or KiD-setup related, we should be able to resolve it and provide a consistent solution for the pattern of precipitation across these two different implementations of SDM.

Such a consensus across [SDMs](#) is something bulk/bin models have been unable to provide and it supports the use of [SDM](#) to model shallow cloud organisation. One reason is that we can use [SDM](#) to explore the effect of knowledge gaps on shallow clouds with more confidence than if we used bulk or bin models that our results are not confounded by high sensitivity to arbitrary model parameters. For example, if we see that in one [SDM](#) without accounting for turbulence there is too little homogeneous mixing at the boundaries of shallow cumuli as compared to observations, then we can say with some certainty that turbulence is important for shallow cloud development and is not just compensating the outcomes of the modelling choices made by a particular [SDM](#) implementation. In contrast there would be less certainty of such a conclusion made using a bulk/bin scheme because it would be compromised by the fact there is no agreed solution for how such models should behave.

Having a numerical benchmark also aids modelling of shallow cumulus organisation because it can help identify where existing large-domain [LES](#) and [GSRMs](#) using bulk schemes need modification. Further work with our [KiD](#) setup could identify the ways in which the assumptions made in a bulk scheme differ from the solution given by [SDM](#) and then elucidate how to adjust the parametrisations in bulk schemes to make them more physical, for example regarding the functional form of rain [DSDs](#). In turn such improvements could change the representation of shallow clouds in large-domains, including their organisation, or at the very least make their representation more physically motivated.

Given that our work therefore supports the use of [SDM](#) to model shallow cumulus organisation, in the next section I explain how we used Cleo to better constrain the amount of, and controls on, rain evaporation in the sub-cloud layer of shallow cumuli.

3.2 CONSTRAINING RAIN EVAPORATION

3.2.1 *A Model For Constraining Rain Evaporation*

Despite the importance of rain evaporation for shallow cumulus organisation there is a scarcity of models and observations. As explained in Section [1.3.1](#), rain evaporation is a central ingredient for shallow cumulus organisation. The vertical profile of evaporative cooling is important for the distribution of latent cooling and moistening and this consequently affects the boundary layer stability locally as well as the triggering of cold pools [e. g. [120](#), [199](#), [236](#), [250](#)]. The evaporation fraction, the fraction of precipitation from cloud-base which reaches the surface, is the other principal metric for rain evaporation because it determines the intensity of surface precipitation. Furthermore in combination with sub-cloud moisture circulations it determines whether

or not clouds disintegrate, which could have repercussions for their subsequent organisation [136, 199, 221, 224]. However it is difficult to quantify rain evaporation, both its cooling rate and fraction, from direct measurements; retrievals do exist but have large measurement uncertainties and have not yet been applied to shallow cumuli [282]. As a replacement for observations, a few modelling studies that were constrained with observations have been conducted [222, 223], but the results lack robustness and generality. Furthermore, they have never accounted for collisions, and especially collisional breakup, despite several studies suggesting they could have a significant impact on rain evaporation through changes to the DSD (Sec. 1.3.1.2).

In the face of such scarcity, we conducted a novel modelling study using SDM to constrain the amount of and controls on rain evaporation in the sub-cloud layer of shallow cumuli. Our main goal was to create an independent model of sub-cloud rain evaporation so that we could compare its results with those of [222, 223] to assess the reliability of our, and their, constraints on rain evaporation. In doing so, we also aimed to provide a “pipeline” for running Cleo in conjunction with observations from Barbados Cloud Observatory (BCO); primarily because with this pipeline we would be able to compare model estimates of rain evaporation more easily and directly with the upcoming rain evaporation retrievals from the Rain and cloud Organization in the Trades using ObseRvations and models (ROTOR) project. On top of these goals however, we also sought to assess for the first time whether shallow cumuli rain evaporation is impacted by droplet collisions, and especially collisional breakup. The paper about this work, of which I am joint-first author together with my former master student Nils Niebaum, is currently under review at *Atmospheric Chemistry and Physics* and can be found in Appendix C.

To simulate rain evaporation in the sub-cloud layer beneath shallow cumuli we setup 1-D column models using observations from EUREC⁴A [263]. We identified 102 raining clouds from in-situ measurements of the DSD at cloud-base and then identified airborne dropsonde measurements in the vicinity of each raining cloud [32, 48, 68, 69]. From these measurements we setup a 1-D column for each raining cloud with thermodynamic profiles fitted from the dropsondes and superdroplets initialised at the top of the column and sampled from fits to the cloud-base DSDs. Superdroplets then evolved freely in the sub-cloud layer until they reached a steady-state, from which we could look at “snapshots” of the evaporation rate and evaporation fraction for each cloud.

We ran four different simulations for each cloud to assess the impact of droplet collisions and collisional breakup on sub-cloud layer rain evaporation. All the setups modelled droplet motion (droplets falling according to their terminal velocity) and evaporation/condensation. The control setup, “EvapOnly” modelled no other microphys-

ics and is comparable to the setup used by Sarkar et al. [222]. “Evap-Coal” additionally modelled collision-coalescence, and “EvapCoalBuRe-many” and “EvapCoalBuRe-few setups” further included collisional breakup and rebound. EvapCoalBuRe-many and EvapCoalBuRe-few represented the upper and lower bounds on the effect of breakup on rain evaporation, by producing very many and very few fragments respectively. This bracketing strategy was used to test the plausible sensitivity of rain evaporation to collisional breakup, whilst not relying on a particular parametrisation which would have large knowledge uncertainty (Sec. 1.1.2.2). Each of the setups discussed here was created using Cleo’s monoids, simply compiling the microphysical process for each simulation with or without one of Cleo’s treatments for the collision algorithm (Sec. 2.2.3).

3.2.2 The Amount of and Controls on Rain Evaporation

3.2.2.1 The Amount

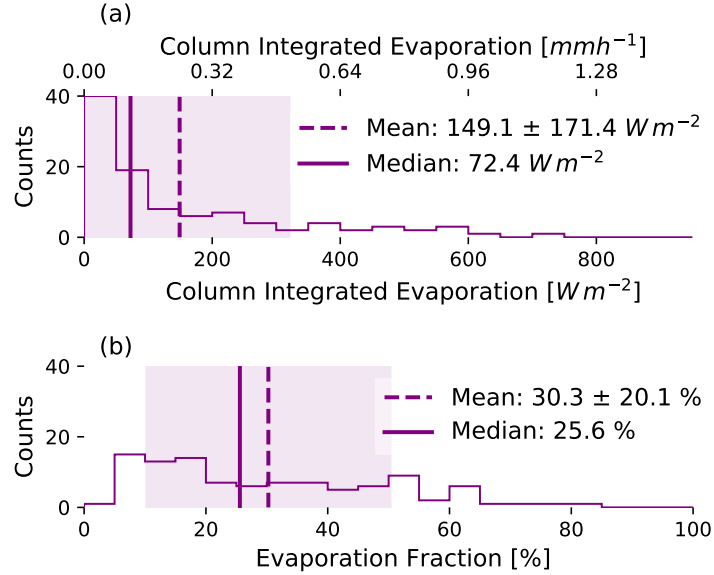


Figure 3.4: Distribution for the 102 clouds in our sample modelled with the EvapOnly setup of (a) column integrated evaporative cooling and (b) evaporation fraction. *This figure is reproduced from C. J. A. Bayley* and N. Niebaum* et al. [40] (App. C Sec. C.4).*

The total sub-cloud layer evaporative cooling from all the 102 clouds in our sample has a skewed distribution ranging from 0 – 800 W m^{-2} with a mean of $150 \pm 170 \text{ W m}^{-2}$ and a median of 72 W m^{-2} ; the evaporation fraction ranges from 0 – 80% and is less skewed, with a mean of $30 \pm 20\%$ and a median of 25.6% (Fig. 3.4). Somewhat surprisingly the total evaporative cooling and the evaporation fraction

display no correlation and, except in a few rare cases, evaporation has a bottom-heavy profile, with its maximum in the lowest third of the rain-shaft. We could explain these findings, and their differences to Sarkar et al. [222], by the factors we found controlling sub-cloud layer evaporation.

3.2.2.2 *The Controls*

Evaporative cooling is mostly determined by the cloud-base Rain Water Content (*RWC*) and the sub-cloud layer relative humidity profile. We find a strong correlation between the cloud-base *RWC* and column integrated evaporative cooling, suggesting that clouds with higher *RWC* have more water available to evaporate in the sub-cloud layer too. For a given *RWC*, a steeper relative humidity gradient induces greater evaporation. However, the relative humidity in the sub-cloud layer is extremely similar between clouds in our sample because the sub-cloud layer is so well mixed, and so we suspect that the importance of the relative humidity profile would be greater if it varied more substantially between clouds or in time.

Evaporation fraction appears to be uncorrelated with evaporative cooling because it depends primarily on the cloud-base *DSD*, regardless of the total *RWC* the distribution holds. *DSDs* with a smaller mean (mass) radius contain more droplets which evaporate efficiently and therefore increase the evaporation fraction. Moreover, we are able to accurately predict the evaporation fraction with a simple analytical single-droplet, mean-field approximation (Fig. 3.5a) (App. C Sec. C.8.1). The exceptions to this approximation are clouds with broad *DSDs* because for those clouds the single-droplet approximation does not summarise the distribution well.

Figure 3.5a also shows the ventilation effect is crucial for determining evaporation, increasing evaporation fraction by almost an order of magnitude in many cases. The ventilation effect is the increase in the evaporation rate due to droplet motion compared to the rate calculated according to Köhler theory for a stationary droplet. It increases the evaporation rate of mm-size droplets by up to a factor of 20 [23], and therefore strongly increases rain evaporation from clouds which have cloud-base *DSDs* with similarly large mean radii. Accounting for the ventilation effect is therefore necessary for correctly simulating cold pool formation, as will be further discussed in Section 3.2.3.

Collisions have negligible effect on rain evaporation in the sub-cloud layer of shallow cumuli. We compared our 1-D column models using Cleo with/without collisions to assess, for the first time, the effect of collisions on rain evaporation. Figure 3.5b shows how accounting for collisions affects evaporation fraction and the same results hold qualitatively for evaporative cooling. Except for cases of severe breakup, the variability amongst the *DSDs* in our cloud sample has a greater effect on evaporation than collisions. Having said that,

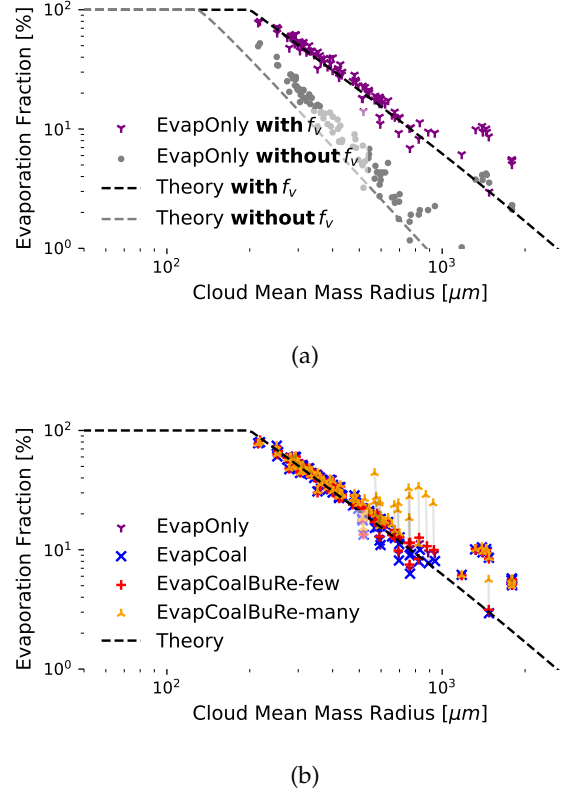


Figure 3.5: Evaporation fraction against cloud-base mean (mass) radius for all the 102 clouds in our sample. “Theory” black dashed lines depict the analytical approximation. (a) The EvapOnly setup demonstrates the importance of the ventilation effect; the grey data shows the results when ventilation is not accounted for. (b) The impact of collisions is negligible; different marker-styles indicate the different setups.

for a handful of clouds with cloud-base rain rates exceeding approximately 3 mm h^{-1} , collisional breakup occurs frequently enough that it can have an appreciable effect on rain evaporation. Therefore, although we find collisions have negligible impact on evaporation from shallow cumuli, they are likely to be significant for more heavily precipitating clouds.

3.2.2.3 Agreement with Previous Studies

Our results are mostly in agreement with Sarkar et al. [222] and Sarkar et al. [223], which used bin microphysics in a o-D model and neglected droplet collisions. Our column integrated evaporative cooling rates are slightly higher than, but still comparable with, Sarkar et al. [222] and although our evaporation profiles are more bottom-heavy and our evaporation fractions are concordantly lower, this can be explained by the fact our model is based on measurements taken further downstream in the Trades than theirs. Further down-

stream, the cloud layer is deeper and hence *RWCs* are slightly larger, and cloud-base *DSDs* have larger droplet radii. Based on our findings this is indeed expected to result in greater evaporative cooling and lower evaporation fraction. We also agree with Sarkar et al. [223] that larger droplet radii decrease evaporation fraction and that environmental conditions play a secondary role to microphysical influences on rain evaporation beneath shallow cumuli. Whilst they identify cloud droplet number as the dominant control on evaporative cooling, this could be somewhat synonymous with what we identify as cloud-base *RWC*, meaning that here too we are in agreement. Since we also find we can neglect droplet collisions, our results strengthen the robustness of the findings of Sarkar et al. [222] and Sarkar et al. [223].

3.2.3 Summary

Our results, in agreement with previous studies [222, 223], emphasise the *DSD*, and therefore the microphysics which determines it, as the dominant control over rain evaporation in the sub-cloud layer beneath shallow cumuli. Correctly representing the cloud-base *DSD* is crucial for predicting evaporation fraction, likewise cloud-base *RWC* is necessary for evaporative cooling. Environmental factors, such as relative humidity and cloud-base height, play only a secondary or insignificant role, most likely because they show such little variation between shallow cumuli. The concordance of these results with previous studies strengthens their reliability and helps to overcome the deficit of direct measurements or radar retrievals for evaporation.

Furthermore, we find that collisions between droplets have a negligible effect on rain evaporation, which bodes well for modelling rain evaporation with simpler models of warm-cloud microphysics, such as bulk microphysics schemes. The analytical approximation which well captures our *SDM* results shows that other simple models should also be able to correctly diagnose rain evaporation rates, at least for the vast majority of lightly precipitating clouds, provided they can accurately calculate cloud water content and the mean radius of the *DSD* (although this is by no means guaranteed and would require validation against observations or, failing that, *SDM*). Crucial however, is that models account for the ventilation effect. Ventilation increases rain evaporation rates by up to a factor of 10 and is therefore a first-order influence on cold pool formation and hence cloud organisation. This result has also been corroborated by *SDM* simulations which did not account for the ventilation effect and therefore struggled to generate strong enough cold pools to simulate squall lines (S. Shima, personal communication, April 22, 2025).

Our column model is ready to be used with any analogous measurements to the dropsondes and cloud-base *DSDs* we have used in

this study. It could therefore be used in conjunction with long-term measurements from [BCO](#), which would make a useful comparison for upcoming radar retrievals from [ROTO_Я](#). Furthermore in this study we used in-situ [DSD](#) measurements from the SAFIRE ATR42 aircraft [32], however many more such measurements were also taken during [EUREC⁴A](#) from aboard the Twin Otter aircraft and these are complemented by radar and aerosol measurements from [BCO](#) [158]. We could therefore easily extend this study using their measurements too. With either [ROTO_Я](#) or more measurements from [EUREC⁴A](#) we might answer whether our results hold for a wider set of environmental and microphysical conditions than our sample allowed, especially more heavily precipitating cases. In such cases we might also explore the relative importance of cloud layer evaporation from outside the edges of sheared clouds too, as well as start to see the impact of collisions between droplets, especially those causing frequent droplet breakup. Using radar-based observations from either [ROTO_Я](#) or [EUREC⁴A](#) naturally also raises questions about how we compare radar measurements with microphysics model output, as will be elaborated on in Section 4.2.2.

Nevertheless, regarding shallow cumuli, we have provided new constraints on the amount of, and controls on, rain evaporation in their sub-cloud layer; rain evaporation which is crucial for triggering cold pools and therefore shallow cumulus organisation.

CONCLUSION

Perhaps cloud photographs can supply the needed observations.

— E. N. Lorenz [160]

4.1 CLOSING THOUGHTS

4.1.1 *Where Have We Come To?*

Our ability to model tropical precipitation and cloud organisation is being improved by [SDM](#). Many of the persistent differences in the macrophysical properties of warm-clouds across [GSRMs](#) and [LES](#), and in comparison to observations, arise from the bulk and bin schemes we conventionally use to model cloud microphysics. These schemes introduce inherent model uncertainty through the modelling choices they make to represent hydrometeors and microphysical processes, and they confound our ability to explore uncertainty stemming from gaps in our knowledge of cloud microphysics. Although [SDM](#) introduces its own, different, modelling choices as well as computational challenges, [SDM](#) offers a more fundamental representation of microphysics which eliminates conventional sources of inherent model uncertainty and makes it easier to understand knowledge gaps. The benefits of [SDM](#) are already being exploited and make it an extremely appealing model to apply to mesoscale (20 – 200 km) domain [LES](#) of shallow cumulus, because it would enable us to improve our representation and understanding of the coupling between cumulus organisation and precipitation mediated by warm-cloud microphysics.

Motivated primarily by the need to make [SDM](#) computationally feasible for elucidating the role of microphysics in mesoscale domains, in the last four years I, together with my collaborators, have created a new computational implementation of [SDM](#) from scratch. We therefore designed our novel [SDM](#), Cleo, with certain characteristics to make it well-suited to large simulations on high-performance computers as well as conducive to microphysical process studies. As a result, Cleo’s computational performance on shared memory architectures is highly promising, with ideal superdroplet-scaling and sufficient strong-scaling to invite Cleo’s next stage of development. This next stage of development will be elaborated on in the following section and primarily regards the particle transport via [MPI](#) on exascale computers. It is the final crucial aspect of making mesoscale domain [LES](#) using [SDM](#) computationally feasible.

In terms of its microphysics, Cleo is a fully-fledged [SDM](#) capable of simulating all the principal microphysical processes in warm-clouds. This includes explicit representation from Köhler theory of droplet (de-/re-)activation, condensation, and evaporation, and an explicit representation of droplet motion including their terminal velocities, as well as the canonical [SDM](#) collision-coalescence algorithm according to Shima et al. [241]. On top of this, we created an extension to this algorithm which can be used to probe the impact of collisional breakup within the bounds of its large knowledge uncertainty (regarding the probability of collisional breakup and fragment size distributions), and in a computationally efficient manner. Simulations using Cleo can be 0-D, 1-D, 2-D, or 3-D and Cleo can be coupled to different dynamical drivers, optionally with an independently determined grid and domain decomposition. Cleo is therefore a novel computational implementation of [SDM](#) which satisfies the necessary requirements for simulating warm-clouds to better understand the role of microphysics in shallow cumulus organisation.

In this dissertation I presented two studies using Cleo as 1-D column models to constrain uncertainties concerning warm-cloud precipitation and rain evaporation, both of which are relevant for shallow cumulus organisation. In the first, we used Cleo in collaboration with another [SDM](#), PySDM, to clarify whether precipitation is sensitive to inherent model uncertainties in [SDM](#) which arise from the chosen method of superdroplet initialisation, parametrisation of collision efficiencies, and number of superdroplets. Encouragingly, these arbitrary modelling choices in [SDM](#) are unlikely to explain significant differences in precipitation. Our work also supports the argument that using [SDM](#) we can provide a numerical benchmark for warm-cloud simulations upon which, for a given set of microphysical processes, all microphysics models should agree. In light of this, we used Cleo in the second study to constrain uncertainty in rain evaporation in the sub-cloud layer of shallow cumuli, in particular exploiting its flexible representation of microphysical processes and knowledge gaps. We found that collisions between droplets have negligible effect on lightly precipitating conditions, whereas the ventilation effect increases rain evaporation by up to a factor of 10. On top of this, our magnitude of and controls on rain evaporation are broadly in agreement with previous work using bin microphysics in a simpler setup [222, 223], which strengthens their reliability. Both the studies summarised here build on our understanding of shallow cloud precipitation and how we can model it. Furthermore they demonstrate some of the benefits of [SDM](#) and Cleo especially; they therefore support Cleo's use in larger domains which can capture the relationship between precipitation and shallow cumulus organisation.

4.1.2 *Where Can We Go From Here?*

There are several fruitful ways in which the two studies presented in this dissertation could be extended. Regarding the [SDM](#) inter-comparison study, the addition of more [SDMs](#), and possibly other Lagrangian microphysics models, would add confidence to our conclusions. So too would more complicated setups, for example the 2-D [KiD](#) [244], or a similar setup in which microphysics can feed back on thermodynamics. As I already explained in Section 3.1.3, we could also improve bulk schemes by comparing their behaviour in our 1-D [KiD](#) to that of [SDM](#). Regarding the second study, our column model using Cleo for rain evaporation is ready to be combined with other observations. We can therefore explore whether our constraints on the controls on rain evaporation hold in more heavily precipitating regimes, where the role of collisional breakup may be greater and the effect of rain evaporation may saturate the relative humidity profile in some cases. As well as this, we could profit by extending the setup from simple “snapshots” of each cloud to include their time evolution, for example to explore the effects of aerosol scavenging and how quickly the sub-cloud layer becomes saturated. The relevance of cloud layer evaporation from the edges of sheared clouds and the effect on rain evaporation of turbulent entrainment in downdraughts could also be better understood with our column model.

The original motivation for Cleo still stands, to use [SDM](#) in [LES](#) with domains large enough to resolve mesoscale cloud organisation. The clear next step towards this goal is performing the first [LES](#) with Cleo fully-coupled to a fluid-dynamical model such as [ICON](#), and we are now in a position to sketch out the feasibility of such [LES](#) in large enough domains. Consider [LES](#) of mesoscale cloud organisation requiring $\mathcal{O}(10^{10})$ superdroplets to have a $150 \text{ km} \times 150 \text{ km} \times 7 \text{ km}$ domain with grid-spacing 150 m horizontally, 40 m vertically, and 128 superdroplets per grid-box. For a conservative estimate of the cost, let us assume that overhead from particle transport via [MPI](#) and communication with the dynamics-solver increase Cleo’s wall-clock time per simulated-second by a factor of 10. Then, by using 250 “Levante-like” [GPU](#)-nodes (1000 NVIDIA A100 [GPUs](#)), the performance tests summarised in Section 2.3.2 indicate that Cleo would take one second of wall-clock time per simulated-second and have an order of magnitude less memory consumption than the maximum on each node. That would make Cleo’s throughput approximately a factor of 10 less than the strong-scaling limit of [ICON](#) [133], which is reasonable for [LES](#) and could be improved by increasing the number of nodes and thereby moving closer towards Cleo’s strong-scaling limit. Thus we are confident that large-enough [LES](#) for mesoscale cloud organisation can be pursued, especially given the growing availability of ever more

powerful exascale computers. We should therefore consider the next stage in Cleo’s computational development towards such large [LES](#).

Being able to use Cleo in domains large enough to resolve shallow cumulus organisation now above all depends on development of its particle transport library. In particular, it depends on improving the efficiency with which superdroplets are communicated via [MPI](#). This problem is common to Lagrangian particle-in-cell methods and relates closely to load balancing, and thus about choosing a suitable [MPI](#) domain decomposition for a given simulation on a given computer architecture. Possible solutions require further investigation, for example exploring dynamic load balancing and/or concealing the time spent on particle transport in asynchronous communication with Cleo’s host dynamical model. Using integer coordinates, as was done by Matsushima et al. [[165](#)], or mixed precision could also offer improvements by reducing the amount of data transfer. Mixed precision is also relevant for reducing the memory demands on each node, as would be selectively reducing the number of superdroplets in certain regions, for example through adaptive resampling of superdroplets [e.g. [233](#), [285](#)] or using Twomey [CCN](#) activation [[78](#)]. The success of such optimisations will determine not just whether large enough domains for shallow cumulus organisation are feasible, but also just how large the number of superdroplets per grid-box and spatial domain could become.

4.2 OUTLOOK

4.2.1 *Extensions Of or Beyond Microphysics*

Given the flexibility we have designed Cleo with, we can easily extend Cleo to include further microphysical processes. The crudest part of Cleo’s microphysics is its representation of collisional breakup. As our knowledge of the probability of breakup and the fragments produced by breakup increases, naturally so too can we incorporate that in the fidelity of Cleo’s parametrisations for them. Furthermore, arguably the biggest deficiencies in Cleo’s representation of warm-rain lie in the various ways in which turbulence could influence microphysics (although this reflects prevalent knowledge gaps in them). To the extent one may want to assess their importance, we could therefore add representations for the effect of subgrid-scale turbulence on supersaturation, collision efficiencies, droplet clustering and droplet velocities [[54](#)]. Of course increasingly detailed microphysics could easily be implemented in Cleo, for example introducing a process to represent subgrid-scale mixing [[100](#)] or cloud electrification [[310](#)], but discernment is required to avoid the “dilemma of growth” which jeopardises understanding [[87](#), [207](#)].

In contrast, the extension of Cleo beyond warm-phase microphysics would provide clear benefits with fewer liabilities. Such extensions are also facilitated by the flexibility in Cleo’s definition of microphysical processes, achieved via Monoids as explained in Chapter 2, and how easily superdroplet attributes could be modified. The most obvious one is the implementation of ice microphysics, as is already under exploration [36, 242, 249, 303]. However, SDM is also reaching further afield in modelling marine plankton [122] and nuclear fallout [168], and Cleo could be easily adapted to other self-interacting particle-based components of our climate system. This would bring with it all Cleo’s computational benefits, for example to modelling icebergs or animal populations. Finally, we could even use Cleo for modelling silicate-based clouds on exoplanets, to explore atmospheres beyond our climate system.

4.2.2 *Observations of Shallow Cloud Organisation*

Recent field campaigns have given us exceptional observations of tropical shallow cumulus. During the two months I spent participating in ORCESTRA [264], I worked as the cloud radar operator aboard one of the aircraft and witnessed the use of lidar and a visible scanning spectrometer aboard too¹. Amongst many other lessons, I began to appreciate the sheer wealth of information hidden in the interaction of droplets with radiation. The spectrometer made highly accurate measurements of DSDs at cloud tops, encapsulating the effect of microphysics combined with cloud-top turbulent mixing [204]. The radar and lidar measurements offer high temporal and spatial resolution which reveal precise details of aerosols, precipitation, ice habits and turbulence, and also offer reasonably large samples of different cloud types, in my opinion remarkably so. Similar statements could be made about EUREC⁴A and BCO radar and lidar measurements [257, 263], and current satellites such as Earth Cloud, Aerosol and Radiation Explorer (EarthCARE) [302] offer a different perspective with orders of magnitude larger spatial and temporal coverage. In combination with all this remote sensing, in-situ measurements from particle spectrometers, disdrometers, particle image velocimetry and holographic instruments make for an extraordinary collection of relevant data.

By performing LES with Cleo in large-enough domains, we would be able to, for the first time, assess shallow cumulus organisation modelled with SDM in comparison with this abundance of observations. The first logical comparison would be extending the work of

¹ Specifically I worked aboard High Altitude Long range research aircraft (HALO) [259] as the HALO Microwave Package (HAMP) [170] operator and witnessed the use of Water Vapor Lidar Experiment in Space (WALEX) [306] and spectrometer of the Munich Aerosol Cloud Scanner (specMACS) [61].

Vial et al. [295] and Schulz and Stevens [232]. These works evaluated the fidelity to EUREC⁴A and BCO observations of shallow cumulus organisation in LES using ICON with a bulk scheme. Amongst other things they compared different cloud patterns' precipitation, cloudiness, and vertical profiles. By extending this work to include LES with Cleo we would gauge the extent to which SDM improves their representation. Furthermore with increasing the number of superdroplets LES with Cleo would converge towards a simulation of fundamental equations for microphysics we understand, and with increasing resolution we would in tandem resolve more turbulent eddies. We would therefore be able to decipher if our current knowledge of cloud microphysics is sufficient to correctly capture shallow cumulus, their precipitation and their organisation, in a way we can not with conventional bulk/bin microphysics models.

Going one step further, rather than simply comparing LES to observations, we could revisit *the way* we perform such comparisons. Radar and lidar simulators are used to convert LES output into observable quantities but they suffer from similar uncertainties to those of conventional microphysics schemes; for example in the case of bulk method simulators they are forced to make assumptions about the functional form of the DSDs, and in the case of bin methods they are limited by the curse of dimensionality. We already know that SDM overcomes these sources of inherent model uncertainty and provides an explicit representation of hydrometeors and aerosols. We could therefore use Cleo to eradicate many of the uncertainties in current radar and lidar simulators by performing radiative transfer calculations on superdroplets in the same fundamental way we would for real droplets or aerosols; and we could even create similar simulators for visible scanning spectrometers. With computing power increasing we will also be able to use Cleo in ever larger domains. In GSRMs the interactions between clouds and circulation primarily result from three processes: phase changes, radiative transfer, and turbulent transport of air parcels [34]. The complementarity of Cleo and GSRMs therefore once again would become apparent because SDM has the potential to derive not just phase changes and the turbulence motion of hydrometeors directly from first principles, but also radiative transfer. In doing so, Cleo would therefore become not just a powerful model for microphysics and cloud organisation, but also for building a stronger relation between our simulations and what we can actually observe. Not necessarily just in the warm-phase and the tropics, but for all clouds in our current and changing climate.

APPENDICES

CLEO: THE FUNDAMENTAL DESIGN FOR HIGH COMPUTATIONAL PERFORMANCE OF A NEW SUPERDROPLET MODEL

The work in this appendix is currently under review at *Geoscientific Model Development* as C. J. A. Bayley et al., ‘CLEO: The Fundamental Design for High Computational Performance of a New Superdroplet Model’, *EGUsphere* **2025, 1–33 (2025).**

Authors: Clara J.A. Bayley^{1,2}, Tobias Kölling¹, Ann Kristin Naumann^{1,3,4}, Raphaela Vogel³, and Bjorn Stevens¹

¹Max-Planck-Institut für Meteorologie, Hamburg, Germany

²International Max Planck Research School on Earth System Modelling, Hamburg, Germany

³Meteorologisches Institut, Universität Hamburg, Hamburg, Germany

⁴Ludwig-Maximilians-Universität München, Munich, Germany

Author contributions: CJAB is the creator of Cleo and main developer, she also wrote and edited the manuscript. TK wrote parts of Cleo’s code and heavily influenced its development through teaching and discussing with CJAB. Most notably TK came up with the idea for how to use C++ concepts to create Cleo’s Monoids, and the adaptive timestepping algorithm; he also introduced using the Zarr storage specification version 2 and ragged arrays for superdroplets for data output. AKN and RV supervised the project, providing direction and support, and teaching about parts of cloud physics relevant to this paper. BS conceptualised the project, oversaw Cleo’s development and had many discussions with CJAB which shaped the writing of the paper and provided scientific and technical support. He also helped analyse the performance results. TK, AKN, RV, and BS all gave extensive feedback which contributed to the writing of the manuscript.

A.1 ABSTRACT

Cleo is a Super-Droplet Model (SDM) designed for performance portability on high performance computer architectures and with the intention of modelling warm-clouds in domains large enough to resolve shallow meso-scale cloud organisation $O(100\text{ km})$. This paper introduces Cleo’s novel C++ implementation of SDM, in particular how we map SDM theory to computations which optimise performance, primarily by conservative memory usage and efficient memory access patterns. To further speed-up simulations and to ensure a portable and maintainable code, we avoid conditional code branching and implement thread parallelism through the Kokkos library. As a result Cleo shows optimal linear scaling with increasing number of super-droplets and can use CPU and GPU thread-parallelisation across a diverse range of computer architectures. But Cleo is not just a model for computational performance, it is also designed for warm-cloud process understanding. Cleo possesses a high degree of flexibility, especially with regard to the configuration of microphysical processes and data output, that makes it well-suited to analysing sensitivity to microphysics. Cleo is therefore a new SDM ready to be used for understanding warm-cloud processes.

A.2 INTRODUCTION

Persistent and large discrepancies between observations of warm-rain and Large Eddy Simulations (LES) have been ascribed to cloud microphysics for decades [e.g. 2, 3, 214, 289] [references in 130, 182]. With the advent of Global Storm Resolving Models (GSRMs) which remove the need to parametrise convection, obscurity in cloud microphysics is now also one of the leading sources of uncertainty in global models [e.g. 17, 140, 176, 185, 262, 268, 272]. To tackle uncertainty in both LES and GSRMs we not only need to address fundamental gaps in our microphysics knowledge, but we must also reconsider how we represent cloud microphysics in models [84].

Conventional bulk and bin microphysics schemes have intrinsic deficiencies, for example due to numerical diffusion, the categorisation of condensates into discrete types, and gross assumptions about particle size distributions. The consequent deficiencies in simulated clouds are well documented and contribute to uncertainties in their radiative properties as well as precipitation [e.g. 121, 130, 232, 265]. To state just one example, numerical diffusion in physical space is known to activate droplets at cloud-edge and broaden the droplet size distribution, and so may cause unphysical precipitation from stratocumulus [57, 95, 181, 254]. Problems like these are inherent to bulk and bin microphysics schemes and therefore cannot be eradicated by refinement of such models. Indeed, decades of research has aimed to limit the impact of numerical diffusion on cloud microphysics, and yet the problem still persists [84]. If we seek to resolve such issues definitively, then the fundamental limitations of bulk and bin schemes need to be surmounted.

The Super-Droplet Model [SDM; 241] [see also 9, 11, 21, 36, 96, 187, 215, 249], is a fundamentally different model of cloud microphysics with different limitations and a number of key conceptual advantages [84]. In SDM the cloud condensate population is modelled by “superdroplets”. Superdroplets’ motion, microphysics and attributes (for example their masses and radii) are similar to what we expect of real condensate particles; however a superdroplet also has a “multiplicity”, which expresses how many real

particles that superdroplet represents. SDM is a much closer representation of the underlying physics than bulk and bin schemes and, because it is a Lagrangian model, it is non-diffusive (in both physical and mass space) [84, 181]. Furthermore, SDM has a profound convergence property because as the number of superdroplets increases, the model tends towards a Direct Numerical Simulation (DNS) [241]. This makes SDM’s physical interpretation both conceptually elegant and straightforward.

SDM is also particularly well-suited to trends in high performance computing. Its coalescence algorithm conserves the number of simulated particles, i.e. superdroplets, which is beneficial for load balancing. Moreover SDM is extremely parallelisable, which makes it ideal for multi-threaded CPU and GPU architectures and explains impressive scaling with increasing number of superdroplets [11, 19, 58, 165]. Its scaling with increasing microphysical complexity is also better than that of conventional models once the number of superdroplet attributes is greater than about four [241]. Both the conceptual and computational advantages of SDM make it a promising tool for better modelling of cloud microphysics [182], and in step with the growth of high performance computers (HPCs) we can now use SDM for more ambitious simulations than ever before.

One emerging application for SDM is to model warm-rain in domains large enough to resolve shallow mesoscale organisation $O(100\text{ km})$. The evaporation of warm-rain drives cold pool formation and is therefore a key player in mesoscale circulations and cloud organisation [18, 235, 236, 298]. Of particular importance in the tropics, mesoscale cloud organisation has major repercussions for the radiative properties of shallow clouds [33, 192] and is highly sensitive to the formation of precipitation [235, 308]. There is also ample evidence that mesoscale cloud organisation modulates warm-rain formation [e.g. 193, 211, 212, 231, 255], and so there are important two-way interactions between the micro- and meso-scale of warm-clouds which, in order to be understood, require an accurate representation of the droplet size distribution, a clear depiction of the microphysical processes at play, and domains large enough to resolve shallow mesoscale circulations. SDM would be an ideal candidate to meet such requirements, except that, presumably limited by computational resources, the largest domains for SDM have not yet exceeded $O(10\text{ km})$ [165, 228, 242] — far smaller than the $O(100\text{ km})$ domains needed for resolving mesoscale organisation, and far smaller than that which may be feasible in the age of exascale computing. This is just one example of how, *if* SDM could be made inexpensive enough at large-scale, it would benefit the community who use large-domain explicitly-resolved-dynamics models — including not just LES, but also GSRMs.

Hence we are creating a SDM called Cleo: a novel implementation of SDM to model clouds through Lagrangian microphysics on exascale computers.

In this first paper we introduce Cleo’s fundamental computational structure and explain how it is intended for high performance on exascale computers as well as warm-cloud process understanding. Notwithstanding SDM’s many advantages in comparison with bulk and bin microphysics, SDM faces the typical challenges of Lagrangian particle models; namely simulations can be memory-intensive and it is time-consuming to transport particles in memory. The particle transport’s performance on exascale computers depends on both its shared- and distributed-memory parallelism, and the distributed-memory parallelism strongly varies not only with Cleo’s fundamental structure, but also with the fluid-flow given by the host dynam-

ical model as well as the computer network topology. In a follow-up paper we will therefore present Cleo’s distributed-memory parallelism (MPI domain decomposition) and evaluate its performance specifically for LES on exascale computers. With regard to Cleo’s fundamental computational structure, the memory and particle transport expenses are prevalent when accessing, sorting, and shuffling superdroplets. Conscious of this, we chose Cleo’s memory layout to economically allocate memory and to optimise memory access patterns for loops over superdroplets whilst compromising on other aspects of performance. Secondly, we’ve made Cleo able to efficiently allocate resources on parallel computer architectures; both through its handling of domain decomposition, which allows SDM and the dynamics to distribute work independently of one-another, as well as through its use of Kokkos for portable thread parallelism [283, 284].

As well as for high performance, Cleo has various features to facilitate warm-cloud process understanding. We took the idea of monoid sets from mathematics and created a computational analogy which enables adaptive time-stepping, avoids conditional code branching, and makes microphysics and data output highly configurable. The flexibility of microphysics makes sensitivity studies easy to conduct and, for data output, it’s easy to target output specifically to ones needs. Also for the purpose of sensitivity studies we made it straightforward to switch Cleo’s grid and couple Cleo to different dynamical drivers. The companion to this paper, Bayley et al. [23], describes the numerical methods for the warm-cloud microphysics and droplet motion which Cleo makes available.

Cleo’s fundamental design is described in Sections A.3, A.4, and A.5, with particular emphasis on how it is intended to make SDM simulations suitable for high performance computers (HPCs). In Section A.3 we describe Cleo’s underlying memory layout and how it is designed for efficient memory management; Section A.4 is for how Cleo enables economical resource allocation, and Section A.5 describes Cleo’s flexibility through monoids. Overviews of Cleo’s time-stepping routine and performance on a single node are found in Sections A.6 and A.7, respectively.

A.3 MEMORY LAYOUT

Efficient memory management is principally achieved by Cleo’s underlying memory layout of superdroplets and grid-boxes (Sections A.3.1 and A.3.2). We store superdroplets within each shared memory space in an ordered and contiguous chunk of memory. This optimises access patterns for loops over superdroplets and avoids needing additional buffers which increase memory consumption. However, this layout comes at the cost of sorting superdroplets and increasing cache-misses as the number of superdroplet attributes increases. We explain the details behind this compromise in Section A.3.3.

A.3.1 Grid-Boxes

“Grid-Boxes” (cells) compose Cleo’s spatial domain. Each grid-box defines a distinct region of space for the SDM collision algorithm and thereby determines which superdroplets may interact with one another. In some sense the volume of each grid-box therefore determines the accuracy of the SDM because as the volume of each grid-box decreases given a certain number

of superdroplets, the resolution of the flow-field for superdroplet motion increases, and the multiplicity of each superdroplet decreases, thus decreasing the errors in the SDM.

Grid-Boxes are structures consisting of three parts as shown in Figure A.1a. The “GBx-index” is a unique immutable identifier, typically a constant unsigned integer; the “state” defines the macrophysical properties of the volume, such as the wind velocity and thermodynamics (temperature, pressure, relative humidity etc.); and the “view” of superdroplets specifies all the superdroplets which occupy the volume at a given time.

The state may change due to microphysical processes, for example condensation lowers the water vapour pressure and increases the temperature, or it may change through information received from the host dynamical driver as explained in Section A.4.3.

The view specifies all the superdroplets at a given time whose spatial coordinates fall within the grid-box’s boundaries. Rather than actually containing the superdroplets, it is an indicator to their location(s) elsewhere in memory. For example, the view could be a linked-list of pointers to individual superdroplets, or it could be two pointers to the start and end of a sub-section of a larger array of superdroplets (as it is in our current implementation).

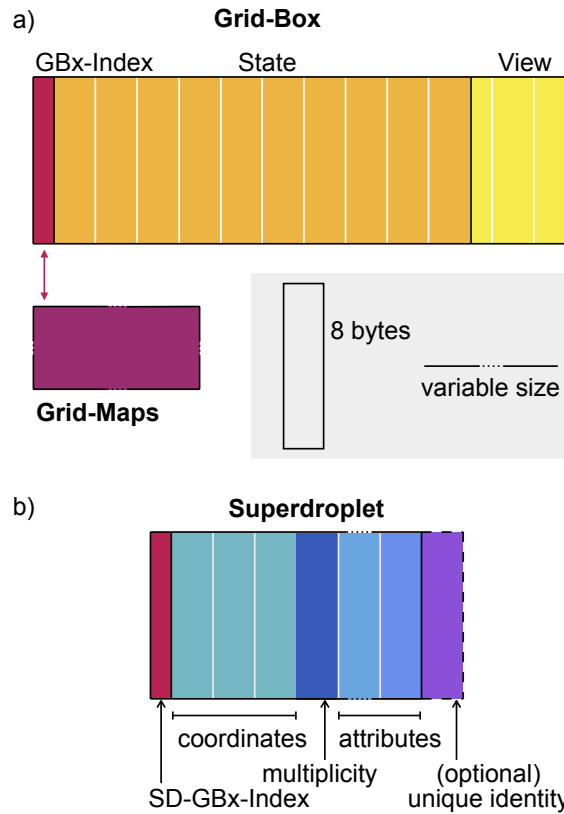


Figure A.1: Schematic for the memory layout of (a) each grid-box, and (b) each superdroplet. The grey box contains a key for the memory consumption of each unit.

Grid-Boxes are agnostic to the underlying grid, meaning they contain no information about their spatial coordinates or their neighbours. This is because LES grids can vary substantially, for example SCALE uses a Cartesian

Arakawa-C grid, whereas ICON uses a icosahedral-triangular Arakawa-C grid [101, 189, 226], and decoupling the microphysics from the grid of the dynamics driver makes it simple to switch between different grids to aid model inter-comparison studies. For a given grid, a collection of maps/functions called “grid-maps” are assembled, to map from a given GBx-index to grid-dependent information. Such maps take a given GBx-index and return for example the coordinates of the boundaries of that grid-box in a particular direction or the GBx-index of the neighbouring grid-box. In doing so they determine the type of grid, the boundary conditions of the model, and whether the model is 0-D, 1-D, 2-D, or 3-D. Thus, changing these properties for different simulations only requires a replacement of the grid-maps. The rest of Cleo, including the grid-boxes themselves, remains unchanged.

A.3.2 Superdroplets

Each superdroplet in Cleo is a structure composed of four, or optionally five parts, as shown in Figure A.1b. The four compulsory parts are its attributes, multiplicity, spatial coordinates, and “SD-GBx-index”. The SD-GBx-index always matches the GBx-index of the grid-box whose boundaries enclose the superdroplet’s spatial coordinates. It is not a requisite of SDM but assists with keeping the superdroplets ordered by grid-box, which is necessary for efficient cache loading as explained in Section A.3.3.

Optionally, each superdroplet can be given a unique identity to facilitate tracking the evolution of individual superdroplets. Alternatively, when tracking is not required, superdroplet identities can be given no unique address. During compilation these identities are optimised out, enabling a simple memory-saving measure especially for simulations involving a very large number of superdroplets. The same optimisation could be applied to grid-boxes’ GBx-indexes (since their positions in memory could act as unique identifiers instead) however we keep GBx-indexes for generality.

A.3.3 Access Patterns

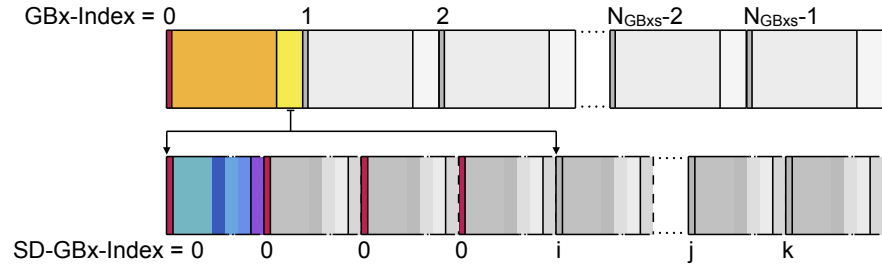


Figure A.2: Schematic for the AoS memory layout of grid-boxes and superdroplets. For illustration here the coordinates of the first four superdroplets are within the first grid-box, hence their SD-GBx-indexes match that grid-box’s GBx-index and that grid-box’s view refers to them.

The memory layout of both grid-boxes and superdroplets in a single shared memory space targets efficient memory access patterns. In this layout, the structures for each grid-box are stored together in a single array (an Array of Structures; AoS), and likewise superdroplets are stored in their

own AoS, in a contiguous block of memory completely separate from grid-boxes. Additionally the superdroplets' array is ordered by increasing SD-GBx-index, and hence the view of superdroplets for each grid-box is formed by references to the start and end of the sub-block of superdroplets whose SD-GBx-indexes match the grid-box's GBx-index. This is all summarised in Figure A.2. The memory layout we have chosen benefits some computational operations at the expense of others. In particular, ordering the superdroplets by SD-GBx-index makes cache loading efficient but requires the additional expense of sorting them during superdroplet motion, and using AoS favours data transfer over Single Instruction, Multiple Data (SIMD) operations. In the following section we discuss in more detail how these conflicts manifest in Cleo's SDM algorithms.

Single, ordered arrays of grid-boxes and superdroplets efficiently handle memory allocations and cache loading for loops over superdroplets. As opposed to having superdroplets scattered in memory, e.g. by having a separate array of superdroplets for each grid-box, having a single array of superdroplets requires less frequent memory (de)allocations when superdroplets move around the domain and therefore also avoids the need for buffers which would increase the memory consumption of the model. The layout we have chosen is also computationally efficient for enacting loops over superdroplets — exactly as occurs during microphysics and while updating superdroplets' spatial coordinates — because both grid-boxes and superdroplets are ordered and accessed sequentially, and the required data from each grid-box is only loaded to the cache memory once for all the superdroplets in a grid-box. We also minimise cache misses during collisions between superdroplets because only superdroplets in the same grid-box may interact with one another and so, by storing them in a contiguous chunk of memory, we maximise the locality of reference between them. As such, Cleo is designed for both efficient memory allocation and cache blocking. Matsushima et al. [165] also employed cache blocking to drastically improve the performance of another implementation of SDM.

Whilst keeping the superdroplet array ordered by SD-GBx-index makes loops over superdroplets highly efficient, it also means that during superdroplet motion we must sort the array, as illustrated in Figure A.3. Superdroplets' spatial coordinates and hence SD-GBx-indexes may change, breaking the crucial condition for cache blocking that superdroplets occupying the same grid-box exist in a contiguous chunk of memory. We therefore apply a counting-sort algorithm to the superdroplet array as the last step of superdroplet motion. Although this algorithm scales linearly with increasing number of superdroplets, it nevertheless involves time-consuming data copies, scatter-gather patterns and/or atomic operations, and it doubles the memory consumption of the superdroplet array. Choosing to keep the superdroplet array ordered is therefore a trade-off. On the one hand it makes loops over superdroplets more efficient, on the other hand it makes superdroplet motion more expensive due to sorting.

Using an AoS for superdroplets rather than a Structure of Arrays (SoA) prioritises efficient data movement over vectorisation. With the AoS layout, irrespective of the number of attributes of each superdroplet, the speed of algorithms which sort/shuffle superdroplets during motion/microphysics scale only with the number of superdroplets. In contrast the SoA layout would need to perform sorting/shuffling on the array for each individual sub-component of the superdroplets separately, resulting in more compu-

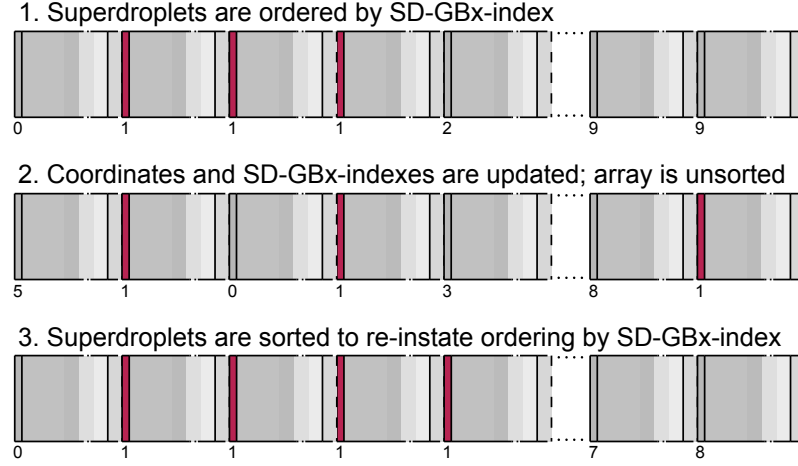


Figure A.3: Schematic of the superdroplet array before, during and after superdroplet motion. To illustrate the ordering, the superdroplets have been given SD-GBx-indexes for grid-boxes from 0 to 9.

tations. Another strength of AoS over SoA is that the addition or removal of superdroplets during motion involves less memory (de)allocations and more efficient cache loading. However, with the AoS layout, operations that act on a single attribute can be less efficient because Single Instruction, Multiple Data (SIMD) parallel processing is harder to apply. Also if the number of attributes of each superdroplet is increased, the number of cache misses likewise increases.

Which data layout results in better performance certainly depends on the computer architecture (e.g. cache size and types of processors) as well as the number of superdroplets and number of superdroplet attributes as a result of the compromise between fast computations through vectorisation and fast memory access patterns. For simulations involving many superdroplets (lots of data transfer) with few attributes (few cache misses), the memory layout we have chosen should be more suitable, but rigorous benchmarking would be needed to confirm the preferred layout for a given setup. In the end, some compromise between the two (an Array of Structures of Arrays, AoSoA) may be the better than either extreme. The decision for the memory layout of Cleo is nevertheless bolstered by the work of Matsushima et al. [165], who also focused on data movement to optimise their SDM's performance, as well as by the rising weight given to the High Performance Conjugate Gradients Benchmark when ranking HPCs, which demonstrates the increasing concern for the performance of memory bandwidth limited computations.

A.4 RESOURCE ALLOCATION

Cleo has the potential to allocate resources economically because of its parallelism and the way it couples to a host dynamical driver. We used Kokkos to implement thread parallelism and Message Passing Interface (MPI) for domain decomposition. As such, Cleo can make use of the available memory hierarchies and execution resources on a diverse set of HPC architectures. To advect thermodynamics Cleo must be coupled to a host dynamical driver capable of advection, however the domain decomposition and computa-

tional resources for the dynamics can be independent from those used by SDM. This allows us greater freedom to improve load balancing and therefore to run economical simulations.

A.4.1 *Kokkos Thread Parallelism*

Cleo depends on Kokkos for performance portable thread parallelism across a diverse range of computer architectures. Kokkos is a collection of libraries which enable us to write a single source code that builds for an arbitrary number of CPU cores and optionally GPU cores [60]. During compilation on a given computer architecture, Kokkos evaluates the hardware available and optional user-defined specifications, to determine how to appropriately allocate the memory and execution resources to achieve high performance. For example, it determines whether to use column- or row-major layouts for arrays of grid-boxes and superdroplets in order to avoid cache misses, and it chooses task sizes for threads of parallelised loops which optimise cache blocking. By doing so, Kokkos implements (hierarchical) parallelism of loops over grid-boxes and superdroplets and selects memory layouts designed to be efficient for a specific processor or accelerator. (Alternatively the memory layouts and task distribution can be manually specified.) The key benefits are that in abstracting the parallelism, we are able to separate the software’s purpose from the details of a computer’s hardware, and we can provide specialised thread parallelism at the same time as portability and maintainability across many architectures. At the time of writing, we have tested CUDA, OpenMP and C++ Threads, and the full list of available parallel execution spaces also includes OpenMPTarget, HIP, SYCL and HPX.

A.4.2 *MPI Domain Decomposition*

For distributed memory parallelism Cleo uses MPI domain decomposition. The layout described in Section A.3 is replicated for the sub-domain given to each MPI process and an extra step in superdroplet motion between the second and third steps of Figure A.3 is activated to send superdroplets between processes. Cleo’s MPI domain decomposition is functioning but we reserve its performance evaluation for a follow-up paper which discusses its optimisation in the context of realistic flows and coupling to a specific dynamical driver.

A.4.3 *Coupling Cleo to a Host Dynamical Driver*

Cleo is designed to run concurrently to a host dynamical driver called a “dynamics-solver” and exchange information with it via a “dynamics-coupler”. In a one-way coupling, Cleo receives information for the state of each grid-box from the dynamics-solver, whereas in a two-way coupling, Cleo also sends state information back. For illustration, a simple one-way coupling is when thermodynamics at a particular time-step are received from a dynamics-solver which stores arrays read from a file. To advect thermodynamic fields as per fluid dynamics, Cleo must be coupled to a dynamics-solver with a fluid-dynamical core.

The dynamics-coupler determines the degree of independence of the grid and of the resource allocation for Cleo and the dynamics-solver. For the max-

imum possible flexibility, Cleo can be coupled to a dynamics-solver through Yet Another Coupler [YAC; 92]. YAC uses MPI communication and can also interpolate variables between different grids meaning that, by using YAC, Cleo and the dynamics-solver can not only have different domain decompositions, but also have different grids. Using different grids is advantageous because it enables the dynamics-solver to compose the domain in the optimal way for its fluid-dynamics. Meanwhile Cleo can compose the domain more favourably for SDM, for example using a nested grid, or grid boundaries which reduce grid-box volumes and/or simplify the numerics of superdroplet motion. In general, using MPI means Cleo and the dynamics-solver do not have shared memory and that their domain decompositions are independent. Whilst this can result in costlier communication, it also maximises our freedom to optimise the load balancing and so economise the allocation of computer resources.

A.5 FLEXIBILITY THROUGH MONOIDS

Cleo uses monoids to enable model flexibility without added run-time from conditional code branching. In abstract mathematics, a monoid is a closed set with an associative binary operation (a semigroup) and an identity element. We apply this idea computationally by creating types, including an identity type, which obey specific constraints to ensure they can be combined pairwise (and associatively) to create a new type which also obeys these constraints. This means any number and permutation of such things can be combined sequentially during compilation to create the instantiation of a template to be used at run-time. We use C++20 concepts to impose the constraints for each monoid on templated types. This has the added benefit of more comprehensible error messages than ordinary template meta-programming.

As an example, consider some code which calls a function “do_microphysics” through an object called “Z”. Z can be any object as long as it obeys some constraints: it contains a function called “do_microphysics”, which takes a view of superdroplets as an argument and potentially modifies it. Now let us create any number of objects, A, B, C, ... which all obey these constraints but have different implementations of the “do_microphysics” function. For example A.do_microphysics may model condensation, whereas B.do_microphysics may model collision-coalescence. We then define an associative binary operation for these objects that returns a new one which likewise satisfies the constraints of the “do_microphysics” function. For example the operation could return “AB” from A and B where AB.do_microphysics = A.do_microphysics followed by B.do_microphysics. Defining the outcome of the operation (AB) to also satisfy the constraints means that we can then combine AB with further objects, C, D, etc. to create one final object, Z, whose “do_microphysics” function is an assembly of all the microphysics we desire. Everywhere else in the code we use Z, for example to call the “do_microphysics” function (rather than A, B, or C etc.), meaning that if we wish to change the microphysics we need only recompile Cleo with a different combination of things comprising Z. This allows us extraordinary freedom regarding the implementation of the “do_microphysics” function whilst keeping the rest of the code intact¹.

¹ Note in general our monoids are not commutative; $AB \neq BA$, condensation enacted before collisions may have a different outcome to condensation after collisions.

Using monoids to enable model flexibility provides faster run-time performance than using conditional code branching. The decisions regarding which functions are called occurs upon instantiation of each templated type during compilation. This is in stark contrast to a code which makes decisions at run-time, for example by using if-statements to determine whether or not to call functions. Indeed, demanding the same degree of flexibility from a code which uses if-statements would require an excessive number of conditional branches — in the example above, to be able to model just A, B, AB, or BA requires three. If this were implemented in Cleo, each branching would add computational expense and reduce code maintainability and readability.

Thus, to ensure run-time performance whilst enabling flexibility, Cleo uses monoids for microphysics and data output. Section A.5.1 details the constraints imposed on objects used to enact microphysics and how such objects can therefore be combined with one another using a binary operator. Likewise, Section A.5.2 details the constraints and associated binary operator for data output.

A.5.1 Microphysical Processes

Microphysics in Cleo is enacted by the instantiation of a templated type which satisfies the constraints of a “microphysical process”. The constraints are chosen to ensure a type instantiated to perform microphysics will function in the time-stepping routine displayed in Figure A.4a. They are that every microphysical process has three (GPU compatible) functions, `next_step`, `on_step`, and `run_step`, each with specific signatures as imposed by the C++20 concept shown in Listing 1. Naturally, types which count as valid microphysical processes may contain additional information beyond these three functions, e.g. a time-step value or a random number generator for a specific microphysics algorithm, however such additional information is not a requirement of a microphysical process.

```

/* define a microphysical process */
template <typename P>
concept MicrophysicalProcess =
    requires(P p, TeamMember &tm, unsigned int t, view_supers supers,
             State &state, NullMonitor mo) {
        { p.next_step(t) } -> std::convertible_to<unsigned int>;
        { p.on_step(t) } -> std::same_as<bool>;
        { p.run_step(tm, t, supers, state, mo) } -> std::same_as<void>;
    };

```

Listing 1: The definition of a microphysical process in Cleo

We also choose to define the associative binary operation $A \oplus B = C$ for types that obey the constraints of a microphysical process such that:

- $C.\text{next_step} = \text{minimum}(A.\text{next_step}, B.\text{next_step}),$
- $C.\text{on_step} = (A.\text{on_step} \text{ or } B.\text{on_step}),$
- $C.\text{run_step} = A.\text{run_step}, \text{ then } B.\text{run_step}.$

This definition enables us to compile a single microphysical process that obeys the time-stepping routine in Figure A.4a but is actually constructed from the sum of any permutation of microphysical processes, as demonstrated in Listing 2. The flow diagram between `C.run_step` and `C.next_step` in Figure A.4a is simply supplanted by calling the `run_step` function of each of the original microphysical processes sequentially, as exemplified by Figure A.4b.

```

/* define how to combine microphysical processes */
template <MicrophysicalProcess M1, MicrophysicalProcess M2>
struct CombinedMicrophysicalProcess {
    M1 a; /**< First microphysical process. */
    M2 b; /**< Second microphysical process. */

    unsigned int next_step(const unsigned int t) const {
        return Kokkos::min(a.next_step(t), b.next_step(t));
    }

    bool on_step(const unsigned int t) const {
        return a.on_step(t) || b.on_step(t);
    }

    void run_step(const TeamMember &tm, const unsigned int t, view_supers supers,
                  State &state, const Monitor auto mo) const {
        a.run_step(tm, t, supers, state, mo);
        b.run_step(tm, t, supers, state, mo);
    }
};

auto operator>>(const MicrophysicalProcess auto a,
                const MicrophysicalProcess auto b) {
    return CombinedMicrophysicalProcess{a, b};
}

// ... ///

/* use a combination of microphysical processes */
const MicrophysicalProcess auto mphys1 = condensation(config);
const MicrophysicalProcess auto mphys2 = collision_coalescence(config);
const MicrophysicalProcess auto mphys3 = collision_breakup(config);
const MicrophysicalProcess auto microphysics =
    mphys3 >> mphys2 >> mphys1; // change this line to change microphysics
}

```

Listing 2: Example code showing how microphysical processes are combined in Cleo. Here we choose “ $\gg$ ” to be the operator in the definition of $A \oplus B = C$ for microphysical processes.

By defining the addition of microphysical processes in this way, we also ensure adaptive time-stepping in which the combined microphysical process always takes the largest possible time-step such that the time-steps of original processes are respected. Consider the microphysical process $C = A \oplus B$. The time-stepping routine in Figure A.4a starts at time=0s, which is less than the final time for time-stepping, t_{final} . When `C.run_step` is first called, both `A.on_step` and `B.on_step` return false and afterwards `C.next_step`

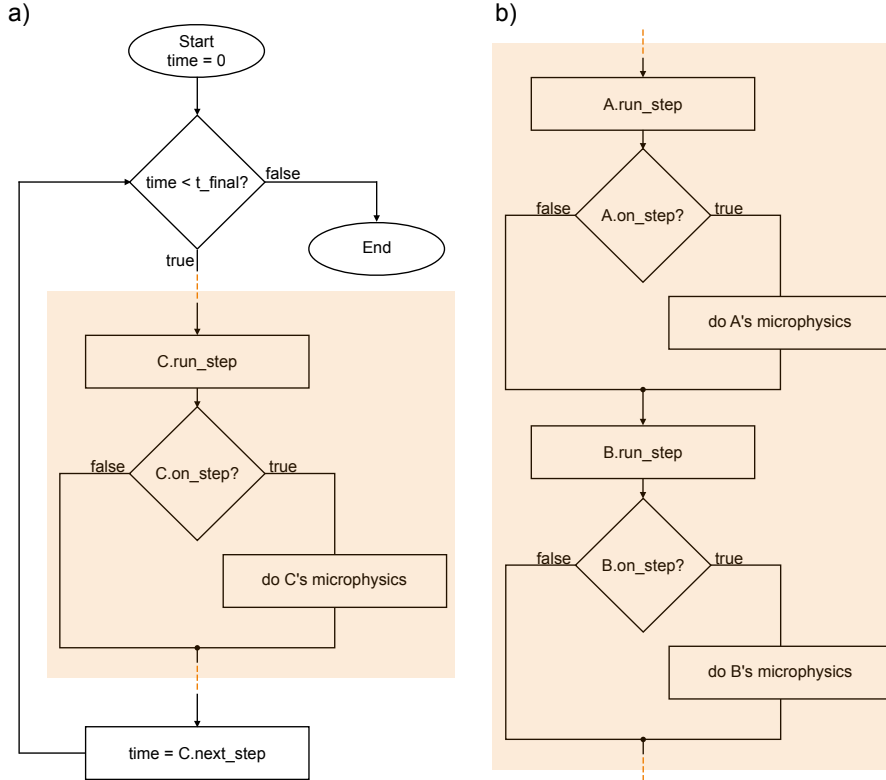


Figure A.4: (a) The flow diagram for the time-stepping routine of some microphysical process “C” as described by the example in the text. (b) The flow diagram replacing that in the orange box of (a) when $C = A \oplus B$.

updates the time to the smaller value out of $A.next_step$ and $B.next_step$. On the next loop iteration, the microphysical process(es) out of A and B which return true from their own `on_step` function enact microphysics, `C.run_step` then returns and time is incremented again. The loop for microphysics enacts this logic repeatedly until t_final is exceeded and the simulation terminates. For demonstration, if $t_final=20s$ and A and B have constant time-steps of 3 s and 5 s, respectively, time advances with the letter in the bracket indicating which microphysics would be enacted at each time-step in the following sequence: 0 s -> 3 s (A) -> 5 s (B) -> 6 s (A) -> 9 s (A) -> 10 s (B) -> 12 s (A) -> 15 s (A and B) -> 18 s (A) -> 20 s (B). In general, A and B need not have constant time-steps and can themselves be formed from the summation of other microphysical processes. Thus we can construct a single microphysical process during compilation which is the sum of any permutation of microphysical processes and enables adaptive time-stepping which adheres to each process’ time-steps.

For completeness of the monoid for microphysical processes, the set has an identity element, `NullMicrophysicalProcess` shown in Listing 3. This is useful as an initial microphysics construction, or in simulations when no microphysics is desired.

```

/* define the identity element for a microphysical process */
struct NullMicrophysicalProcess {
    unsigned int next_step(const unsigned int t) const {
        return std::numeric_limits<unsigned int>::max();
    }

    bool on_step(const unsigned int t) const { return false; }

    void run_step(const TeamMember &tm, const unsigned int t, view_supers
        ↪ supers,
                State &state, const Monitor auto mo) const {}
};

```

Listing 3: The identity element of the microphysical process monoid in Cleo.

A.5.2 Observers

Data output, for example writing information to non-volatile memory or statements to a terminal window, is entirely analogous to microphysics. Output is done by the instantiation of a templated type which satisfies the constraints of an “observer” - a concept for things which have the possibility to read and copy information from Cleo at selected points during time-stepping. Structures called “Monitors” are used by some observers for output which cannot be obtained at the start or end of a time-step, but rather must be monitored during a SDM time-step — for example the mass change due to condensation/evaporation. Monitors are also a monoid in Cleo, but we omit the set’s definition here for the sake of brevity. The constraints on an observer are chosen to ensure the type instantiated as the observer will function in Cleo’s time-stepping routine shown in Figure A.9 (pink boxes). They are that every observer has six functions, `before_timestepping`, `after_timestepping`, `next_step`, `on_step`, `at_start_step`, and `get_monitor`, each with specific signatures as imposed by the C++20 concept shown in Listing 4.

Any type which satisfies these constraints is a valid element of the monoid for observers and can therefore be combined with another element using the associative binary operation $A \oplus B = C$, which we choose to define as:

- `C.before_timestepping = A.before_timestepping`, then `B.before_timestepping`,
- `C.after_timestepping = A.after_timestepping`, then `B.after_timestepping`,
- `C.next_step = minimum(A.next_step, B.next_step)`,
- `C.on_step = (A.on_step or B.on_step)`,
- `C.at_start_step = A.at_start_step`, then `B.at_start_step`,
- `C.get_monitor = A.get_monitor  $\oplus$  B.get_monitor`,

as demonstrated in Listing 6.

```

/* define an Observer */
template <typename Obs>
concept Observer = requires(Obs obs, unsigned int t, view_constgbx gbxs,
                             view_constsupers supers) {
    { obs.before_timestepping(gbxs) } -> std::same_as<void>;
    { obs.after_timestepping() } -> std::same_as<void>;
    { obs.next_step(t) } -> std::convertible_to<unsigned int>;
    { obs.on_step(t) } -> std::same_as<bool>;
    { obs.at_start_step(t, gbxs, supers) } -> std::same_as<void>;
    { obs.get_monitor() }; // should return a Monitor type
};

```

Listing 4: The definition of an observer in Cleo

Exactly like, but independent of, microphysical processes, the sum of different observers ensures adaptive time-stepping. Likewise this means observers can have different output frequencies and ones that are not necessarily factors of each other. In fact, the constraints on the time-stepping functions for observers and microphysics are so similar they could stem from the same parent time-stepping concept, but for simplicity we keep them distinct.

Observers that write data to non-volatile memory are designed to write large datasets efficiently. Buffers hold output data in RAM until they reach a specified chunk size (~10 MB by default). These are then written to chunks of arrays in a dataset according to the Zarr storage specification version 2. Zarr is chosen because it is well-suited to high throughput of large N-dimensional arrays, especially because it enables parallelised reading/writing. The dataset is compliant with the requirements of Xarray, and is therefore also compatible with NetCDF. Data can be written as any data-type of the NetCDF Climate and Forecast (CF) Metadata Convention, but to save memory it is cast to 4-byte types by default. Additionally, arrays of superdroplet data (e.g. their spatial coordinates) are output according to the CF Metadata Convention’s contiguous ragged representation. This is because the number of superdroplets may change over time if superdroplets leave or enter the domain and so the length of data at each output time-step may change. A ragged array, as opposed to a fixed shape array with dimensions of the largest length of data, does not contain empty data points and so saves memory.

For completeness, the monoid for observers has an identity element, `NullObserver`, shown in Listing 5. This is useful as an initial observer construction, or in simulations which do not require output from Cleo.

A.6 TIME-STEPPING CLEO

Cleo’s entire time-stepping routine is shown in Figure A.9. In summary, one time-step consists of seven key stages:

1. Determine the next time-step, `t_next`.
2. If on a coupling time-step, receive thermodynamics and wind velocity fields for SDM from the dynamics-solver via the dynamics-coupler.

```

/* define the identity element for an Observer */
struct NullObserver {
    void before_timestepping(const view_constgbx gbxs) const {}

    void after_timestepping() const {}

    unsigned int next_step(const unsigned int t) const {
        return std::numeric_limits<unsigned int>::max();
    }

    bool on_step(const unsigned int t) const { return false; }

    void at_start_step(const unsigned int t, const view_constgbx gbxs,
                      const view_constsupers supers) const {}

    Monitor auto get_monitor() const { return NullMonitor{}; }
};

```

Listing 5: The identity element of the observer monoid in Cleo.

3. If on an observation time-step, perform observation.
4. Time-step the dynamics-solver from the current time to t_{next} .
5. Time-step SDM from the current time to t_{next} via the sub-time-stepping routine shown in Figure A.9b, enacting microphysics followed by superdroplet motion.
6. If on a coupling time-step, send thermodynamics from SDM to the dynamics-solver via the dynamics-coupler.
7. Increment the time to t_{next} .

The length of one time-step and the exact implementation of each of these parts is determined by types instantiated for the dynamics-solver, dynamics-coupler, grid-maps, observer, superdroplet motion, and microphysical process. For example, when using a one-way coupling the dynamics-coupler's function for step 6 is empty. The sub-time-stepping routine for SDM shown in Figure A.9b is simply an extension of the logic from Figure A.4a to include observers and superdroplet motion as well as microphysics. Although currently it is performed afterwards, to reduce run-time the entire sub-time-stepping routine could in principle be made asynchronous to the time-step of the dynamics-solver.

A.7 PERFORMANCE

Here we show the performance of Cleo's superdroplet-scaling and strong-scaling using OpenMP, C++Threads and CUDA parallelism. The simulations are performed on a single node of Levante HPC at the German Climate Computing Center (DKRZ) using 1, 16, 64 or 128 CPU threads and,

Grid-Boxes	Total Superdroplets	Δx / m	Δy / m	Δz / m
1	2.560×10^2	6000.0	300.0	1500.0
8	2.048×10^3	1500.0	150.0	1500.0
64	1.638×10^4	750.0	75.0	750.0
512	1.311×10^5	375.0	37.5	375.0
2048	5.243×10^5	187.5	37.5	187.5
4096	1.049×10^6	187.5	18.8	187.5
16384	4.194×10^6	93.8	18.8	93.8
32768	8.389×10^6	93.8	9.4	93.8
131072	3.355×10^7	46.9	9.4	46.9

Table A.1: The problem size for the performance tests in terms of the number of grid-boxes composing the domain and the consequent total number of superdroplets and domain grid-spacing.

when CUDA is enabled, an additional GPU². The node’s configuration and the software used are listed in Table A.4. To change the problem size we decrease the volume of each grid-box, holding the domain volume and the number of superdroplets per grid-box fixed [165] [as opposed to fixing the grid-box volumes and increasing the number of superdroplets per grid-box; 58]. This not only decreases the variance in the number of droplet collisions and increases the precision of the droplet size distribution, but also increases the resolution of the flow field and the accuracy of each superdroplet’s position [23]. The better this superdroplet-scaling, i.e. grid-box scaling, the more efficiently we can run accurate simulations.

The simulations are of a 3-D Cartesian domain $6000 \text{ m} \times 300 \text{ m} \times 1500 \text{ m}$ with 256 superdroplets per grid-box, varying grid-spacing as listed in Table A.1 and with the initial conditions shown in Figure A.5a-d. They include condensation, evaporation, collision-coalescence and droplet motion using the configuration listed in Table A.5 for the physics described in Bayley et al. [23]. The droplets in every grid-box are initialised from a bimodal log-normal dry-aerosol distribution of NaCl. The modes are at 20 nm and 150 nm with standard deviations of 1.4 and 1.6, respectively, their relative probability is 3:2, and the total droplet number concentration is 500 cm^{-3} . Superdroplets are created by randomly sampling 256 evenly-in-log-space bins between 5 nm and $1 \mu\text{m}$ — as in the “Single-SIP-init” method of Unterstrasser et al. [286], but in radius not mass space. Superdroplets’ spatial coordinates are uniformly randomly sampled from the space within each grid-box. The dynamics are a simplified version of the 2-D kinematic flow model from Arabas et al. [11]. We prescribe the same divergence-free wind field and hydrostatic equilibrium, but the kinetics are held constant in time, without feedback from microphysics, rather than relaxed towards the initial profiles. To mimic sub-cloud and in-cloud conditions, the temperature has a lapse rate of 9.8 K km^{-1} below 750 m and 6.5 K km^{-1} above, while the

² Tests with CUDA parallelism also used OpenMP for parallelised CPU code, e.g. some parts of initialisation. Since the majority of the time is spent on GPUs, changing the number of CPU threads has negligible effect on the CUDA build’s speed-up and so we only show plots for the tests with 128 CPU threads.

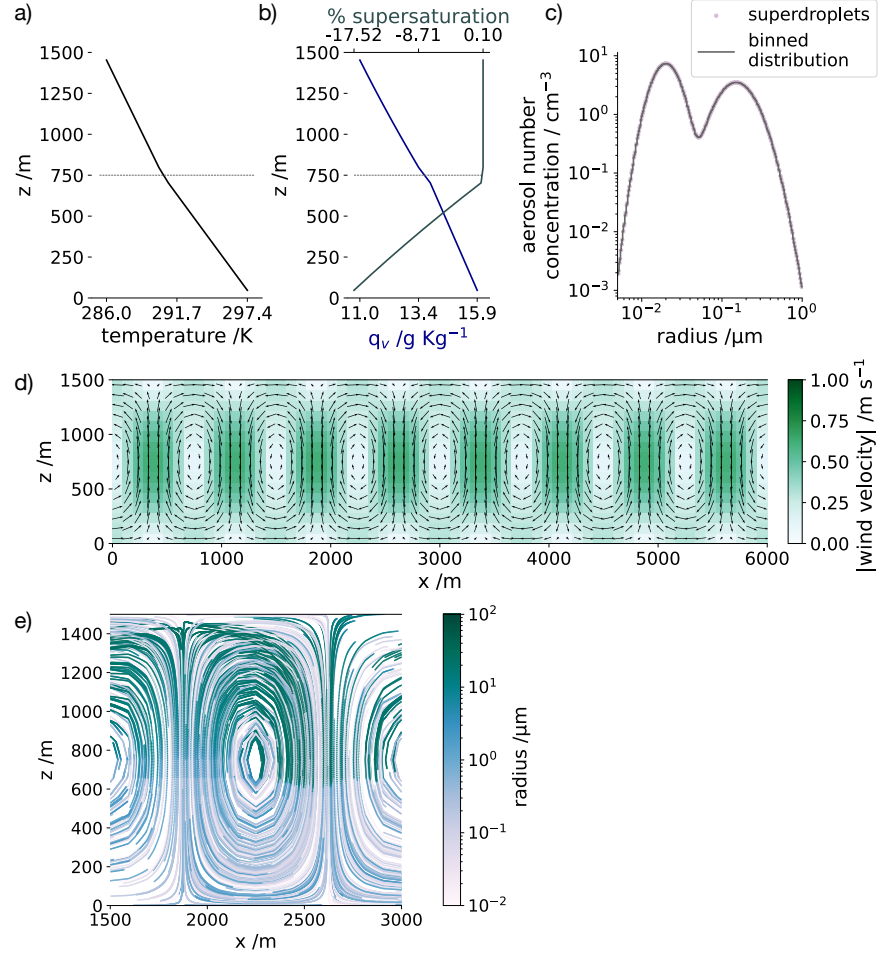


Figure A.5: An example of the performance tests with 16384 grid-boxes. (a-d) The initial conditions, (e) The growth and movement of a random sample of 500 superdroplets between $1500 \text{ m} < x < 3000 \text{ m}$ for the test case with 1.638×10^4 superdroplets overall.

water vapour has a lapse rate of $2.97 \text{ g kg}^{-1} \text{ km}^{-1}$ below 750 m and supersaturation is fixed at 0.1% above. The surface pressure is 1013.15 hPa, the surface temperature is 297.9 K, and the surface water vapour mass mixing ratio is 16 g kg^{-1} . We ran the simulations for 80 mins without data output and present results per 1 s of simulated time. An example of the superdroplets' evolution is shown in Figure A.5e.

The superdroplet-scaling of the wall-clock time is shown in Figure A.6 for the maximal set of resources for each build: serial (1 thread), CUDA (128 CPU threads and 1 GPU), OpenMP (128 threads) and C++Threads (128 threads). The wall-clock time for all the builds is almost entirely time-stepping SDM, as opposed to initialisation or the dynamics-solver, and time-stepping SDM scales linearly with increasing number of superdroplets once the total number of superdroplets is large enough. This is because the bottleneck of microphysics is the (serial) random shuffling of superdroplets during collisions and the bottleneck of motion is the aforementioned counting-sort algorithm (Section A.3.3). Both of these algorithms scale linearly with increasing number of superdroplets. For C++Threads and OpenMP builds,

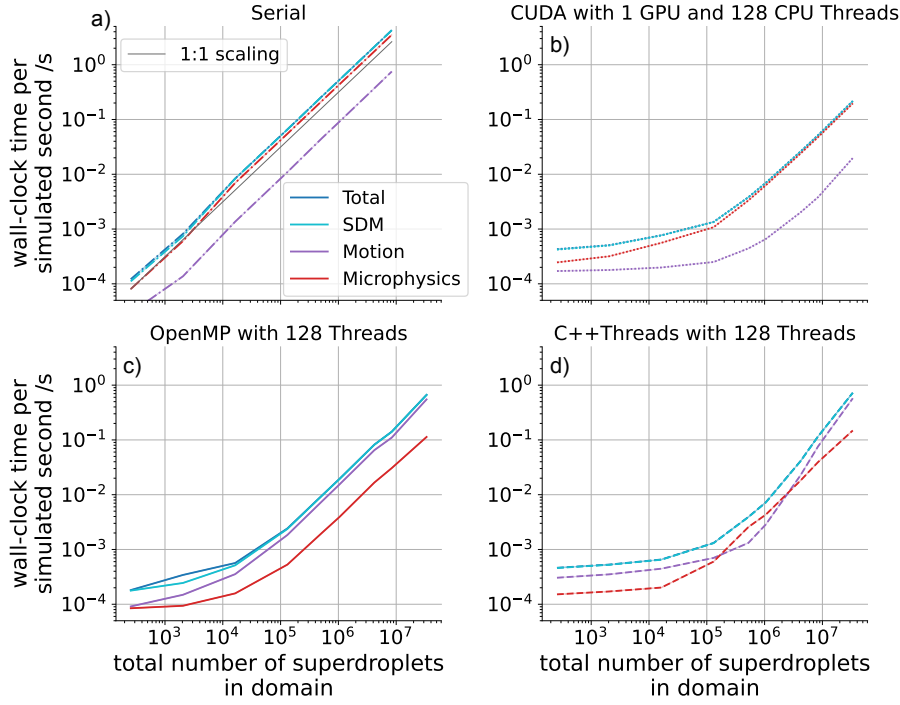


Figure A.6: The superdroplet-scaling of the wall-clock time for each simulated second given the maximum set of resources for each build on Levante HPC: (a) Serial (dash-dotted), (b) CUDA (dotted), (c) OpenMP (solid), and (d) C++Threads (dashed). The colours decompose the total time (blue) into the time spent on SDM (cyan), motion (purple) and microphysics (red)

motion (i.e. sorting) is the dominant bottleneck, while for serial and CUDA builds, it is the microphysics (i.e. shuffling). The time-per-call of motion is, however, three times greater than shown here since the time-step was 3 s. The expense of motion per-call is therefore comparable to that of microphysics for all builds, and considerably more for OpenMP and C++Threads.

The strong-scaling in Figure A.7 shows the speed-up of the wall-clock time with increasing number of CPU threads for various problem sizes². Overall, none of the builds reach optimal efficiency (Figure A.7a); CUDA suffers from microphysics (Figure A.7b), while OpenMP and C++Threads suffer from motion (Figure A.7c).

The limit for CUDA parallelism due to microphysics can be seen in Table A.2. The speed-up of motion maximizes at 150 for the largest problem sizes, however microphysics appears to plateau at smaller problem sizes and at 70. The bottleneck is again the shuffling algorithm we use during collision-coalescence, presumably due to the relatively slow clock-speed of a single GPU thread, and so a different (parallelised) shuffling algorithm may improve performance. Alternatively, the shuffling could be performed concurrently with other parts of the time-stepping routine. For OpenMP, the strong-scaling of microphysics is almost perfect, at 100% efficiency for large enough problem sizes, however, the overall speed-up is limited by the speed-up of motion which is no greater than 8. The scaling of C++Threads is better than OpenMP despite slightly less-optimal scaling of the microphysics because motion scales better for all but the largest problem size. The bottle-

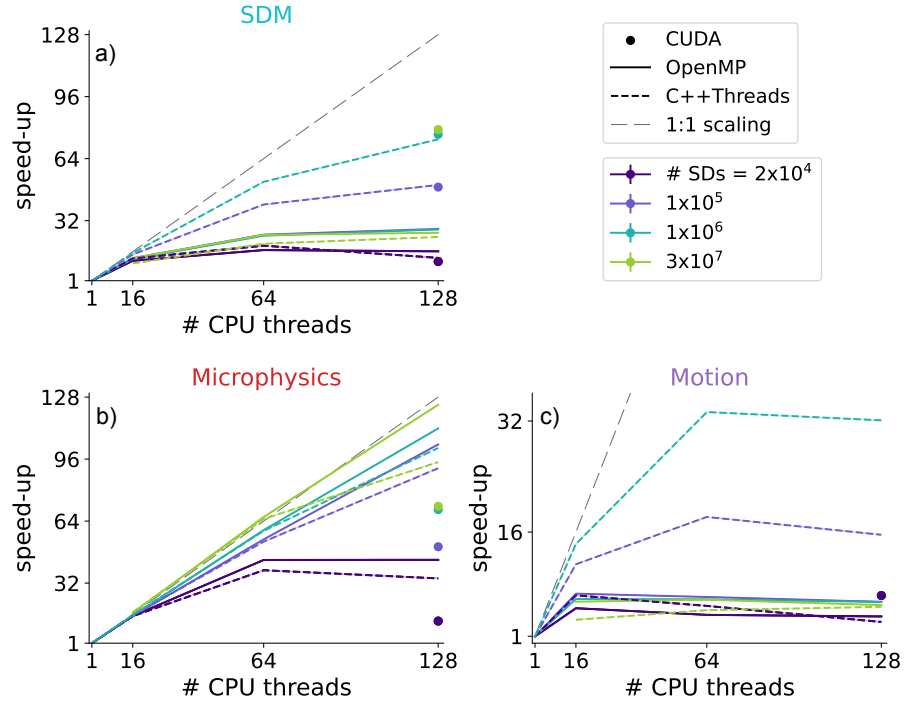


Figure A.7: The strong-scaling of the speed-up of wall-clock time relative to the serial simulations with increasing number of CPU threads on Levante HPC for (a) time spent on SDM, (b) microphysics, and (c) motion. The line-style denotes the build and the colours from purple to green indicate increasing problem size in terms of the total number of superdroplets in the domain. The speed-up of motion from CUDA exceeds the y-axis scale and is reported in Table A.2 instead.

neck of motion on CPUs is not the data copying itself, but rather the synchronisation of threads required by scatter-gather patterns in the counting-sort algorithm. Data copying nevertheless doubles the memory footprint of the superdroplets. A more specialised sorting algorithm which exchanges rather than copies superdroplets and uses the relationship between the GBx-indexes to avoid scatter-gather patterns would therefore provide better performance both in terms of memory usage and speed. Our results are in agreement with Matsushima et al. [165] who found that after optimising their SDM, superdroplet tracking (movement and sorting) was indeed the bottleneck of SDM on CPUs, and for the largest simulation this was predominately due to superdroplet sorting.

The bottlenecks in Cleo’s microphysics and motion are expected to become less acute with more complex simulations. The addition of more microphysical processes, for example to model frozen condensates, will increase the cost of microphysics in ways that decrease importance of the shuffling algorithm. Coupling Cleo to a dynamics-solver capable of advection will diminish SDM’s role in the total run-time and introduce other dynamics-related bottlenecks. Above all, the expense of transporting superdroplets between MPI processes is expected to make the cost of sorting during superdroplet motion inferior to the cost of MPI domain decomposition, even once optimised.

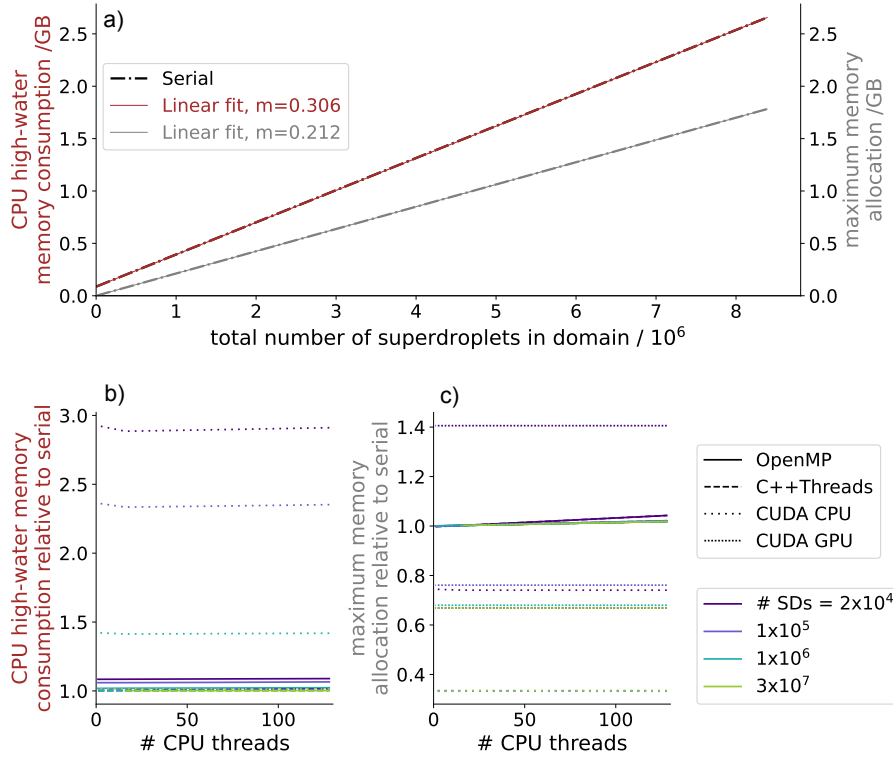


Figure A.8: (a) The superdroplet-scaling of the serial build for the resident set size high-water memory consumption (brown) and for the maximum memory allocation (grey); (b and c) The strong-scaling of the same memory footprints for the parallelised builds: OpenMP (solid), C++Threads (dashed) and CUDA (dotted). The colours from purple to green indicate increasing problem size in terms of the total number of superdroplets in the domain.

Figure A.8 shows the superdroplet- and strong-scaling of Cleo's memory consumption is as expected. The memory footprints increase linearly with increasing number of superdroplets and are almost constant with increasing CPU threads. The overhead from using OpenMP or C++Threads parallelism is almost insignificant. It is larger for smaller problem sizes but still no greater than 1%, and is predominately from the creation of thread-safe random number generators and arrays for the scatter-gather patterns. Due to the large number of threads on a GPU, the overhead of CUDA parallelism for problem sizes less than approximately 10^6 superdroplets can be relatively large because thread-safe random number generators are more expensive than the superdroplets and grid-boxes themselves. For large enough problem sizes however, the overhead of CUDA is also insignificant, and the maximum memory allocation (i.e. the superdroplet memory allocation) can actually be substantially smaller than in serial. The high-water memory consumption, that is the largest resident set size memory used, can be divided into the apparent memory consumed by a single superdroplet and grid-box. As demonstrated in Table A.3, for all builds both superdroplets and grid-boxes are effectively $O(100 \text{ bytes})$, whereas by default stand-alone they are 68 bytes and 120 bytes, respectively.

#SDs in Domain	CUDA Speed-Up		
	SDM	Microphysics	Motion
1.638×10^4	10.9	12.4	6.86
1.311×10^5	49.3	50.8	43.8
1.049×10^6	76.5	69.9	139.5
3.355×10^7	79.0	71.6	152.3

Table A.2: The speed-up of wall-clock time for the CUDA build relative to the serial simulations for problem sizes plotted in Figure A.7.

	Serial	OpenMP	C++Threads	CUDA
% of high-water memory consumption				
Superdroplets	99.3	97.5	97.5	99.1
Grid-Boxes	0.2	0.2	0.2	0.6
Apparent size / bytes				
Superdroplet	314.4	309.8	308.6	332.9
Grid-Box	162.1	162.7	162.0	515.9

Table A.3: The apparent size of superdroplets and grid-boxes calculated from their contribution to the high-water memory consumption for the simulations with 32768 grid-boxes (256 superdroplets per grid-box).

A.8 CONCLUSIONS

To satisfy our curiosity about cloud responses to climate change, SDM offers a promising new way to model cloud microphysics. SDM overcomes many of the problematic features of conventional microphysics representations and this has already proven fruitful, for example in studying the role of turbulence during droplet growth [44, 86, 151]. However, the computational cost currently inhibits SDM from certain areas of research, an important case being the study of interactions between cloud microphysics and their mesoscale organisation. This research requires LES that can clearly depict warm-cloud microphysical processes and resolve droplet sizes whilst also capturing mesoscale circulations. We therefore need a SDM for studying warm-clouds in LES with hectometre resolution and $O(100 \text{ km})$ domains. Cleo is a new computational implementation of SDM intending to be fit for that purpose. Of course, to make such large-domain LES feasible, Cleo must be able to harness both the shared- and distributed-memory parallelism of exascale computers, and whilst the shared-memory performance depends mostly on the fundamental structure of the code, the distributed-memory parallelism is also sensitive to the fluid-flow given by the host dynamical model and the computer network topology. As such, we will evaluate Cleo's distributed-memory parallelism in a follow-up paper specific to LES on exascale computers, whilst this paper discusses Cleo's fundamental com-

putational design and its performance on shared-memory architectures. In Bayley et al. [23], the numerics of Cleo’s warm-cloud microphysics is described.

Efficient memory management is a principal feature of Cleo’s design for high computational performance of large simulations. We have chosen the layout of grid-boxes and superdroplets in contiguous ordered chunks of memory to try to minimise cache loading and support data locality. Whilst keeping superdroplets ordered requires a costly sorting algorithm, this expense will become less significant with larger simulations involving MPI domain decomposition, more complex microphysics and a fluid-dynamical core. The memory layout we have chosen also favours data transfer over SIMD operations, as is deemed appropriate for simulations involving frequent transport of a large number of superdroplets which have a low number of attributes. Since simulations of many superdroplets are exceptionally memory demanding, Cleo’s optional features, such as superdroplet tracking, can be optimised out of simulations. Data output is also designed economically by following the Zarr storage specification to write binary files and by using ragged arrays to represent superdroplet data.

Cleo’s resource allocation is designed flexibly so that large simulations can be economised. To exploit parallelism on HPCs, we use MPI for domain decomposition and Kokkos for abstracted and hierarchical thread parallelism. Kokkos makes Cleo concise and maintainable, but most importantly performance portable, meaning Cleo can make use of the various forms of thread-parallelism available on a diverse set of computer architectures, including those with GPUs. Further flexibility is built into how Cleo couples to a host dynamical driver. The coupling allows SDM and the dynamics to have independent domain decompositions, meaning we have greater freedom to allocate computational resources in a way that achieves load balancing and therefore produces more economical simulations.

Cleo is highly configurable as well as computationally performant. Monoids, defined through template meta-programming with C++20 concepts, enable adaptive time-stepping and make choices regarding which variables to output and which microphysical processes to include highly flexible. In contrast with flexibility obtained through conditional code branching, flexibility via monoids has faster run-time and makes the code more readable. The coupling between Cleo and a host dynamical driver not only allows different domain decompositions, but also makes it easy to switch host model to conduct sensitivity studies and optimise the numerical methods of SDM and the dynamics-solver separately. Additionally, Cleo has flexibility over the choice of grid for SDM. All this flexibility makes Cleo well-suited for detailed study of microphysics, for example evaporation, its influence on hydrometeor evolution and atmospheric dynamics.

Cleo’s current state of development raises opportunity for further optimisation and larger simulations. The bottlenecks of Cleo’s SDM algorithms on a single node are the shuffling during collisions between droplets (acute in GPU simulations), and the sorting during superdroplet motion (in CPU-only simulations). Using a parallelised shuffling algorithm may be more optimal for GPUs, but optimisation could also come from making parts of the shuffling algorithm execute concurrently with other computations at each time-step. Similar optimisations could provide better performance of the sorting algorithm on CPUs, as well as lower its memory consumption. However, whether such optimisations are worthwhile should be assessed with more

ambitious simulations, where the transport of superdroplets between nodes, additional microphysics, and/or the expense of the dynamics-solver and communication with it, could surpass these algorithms as the major bottlenecks. Given Cleo’s current state of development however, we can speculate about the minimum cost of such ambitious simulations. Consider LES with a $150\text{ km} \times 150\text{ km} \times 7\text{ km}$ domain with grid-spacing 150 m horizontally, 40 m vertically and 128 superdroplets per grid-box. In other words, $O(10^{10})$ superdroplets overall. For a conservative estimate on the cost, let us assume that the wall-clock time for superdroplet motion increases by a factor of 10 due to the transport of superdroplets via MPI, but that all of the run-time of the dynamics-solver and communication can be made asynchronous. Then, by using 1000 NVIDIA A100 GPUs (250 “Levante-like” GPU-nodes), the SDM setup from the performance tests presented here suggest that $O(10^7)$ superdroplets per MPI process would need approximately one second of wall-clock time for every second of simulated time. Even for a more conservative estimate, where the performance is slower by a factor of 1000, a simulation of several hours that takes several weeks to run is still conceivable. Future development of Cleo is therefore a promising avenue to pursue towards using SDM to study the interactions between cloud microphysics and mesoscale cloud organisation.

A.9 CODE AVAILABILITY

Cleo is published on it’s GitHub page: github.com/yoctoyotta1024/Cleo, alongside it’s documentation: <https://yoctoyotta1024.github.io/Cleo>. Version v0.39.0 is described and tested in this paper.

A.10 CODE AND DATA AVAILABILITY

All the code and results included this paper can be accessed from the dataset: Bayley [25].

A.11 COMPETING INTERESTS

The contact author has declared that none of the authors has any competing interests.

A.12 ACKNOWLEDGEMENTS

C. J.A. Bayley thanks Shin-ichiro Shima (University of Hyogo, Japan) for his dedicated supervision, constructive feedback on the initial version of the manuscript, and for teaching about SDM. We gratefully acknowledge code contributions to Cleo from Sergey Kosukhin and Lukas Kluft (Max Planck Institute for Meteorology, Germany; MPI-M). A special thanks is given to Yvonne Schrader (MPI-M) for her excellent advice on the graphic designs in this paper, as well as Marco Giorgetta (MPI-M) for conducting the MPI-M internal review.

A. K. Naumann and R. Vogel have received funding which supported this work from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy - EXC 2037 "Climate, Climatic Change, and Society" (project number 390683824). R. Vogel

further acknowledges support from an ERC starting grant (ROTOЯ, grant no. 101116282). This project has received funding from Horizon Europe programme under Grant Agreement No 101137680 via project CERTAINTY (Cloud-aERosol inTeractions & their impActs IN The earth sYstem). The authors further express their appreciation for the work of the developers of the free and open-source software which underlies Cleo, especially from the developers of Git, GitHub, Python, the C++ standard libraries, and, above all, Kokkos. We also thank the Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG) from the information and communication services Cleo's development has benefited from, and finally we thank the Deutsche Klimarechenzentrum (DKRZ) for the computer facilities from project 1183 we used to conduct this work.

```

/* define how to combine observers */
template <Observer Obs1, Observer Obs2, Monitor Mo>
struct CombinedObserver {
private:
    Obs1 a; /**< First Observer. */
    Obs2 b; /**< Second Observer. */
    Mo mo; /**< Combination of First and Second Observers' Monitors */

public:
    void before_timestepping(const view_constgbx gbxs) const {
        a.before_timestepping(gbxs);
        b.before_timestepping(gbxs);
    }

    void after_timestepping() const {
        a.after_timestepping();
        b.after_timestepping();
    }

    unsigned int next_step(const unsigned int t) const {
        return Kokkos::min(a.next_step(t), b.next_step(t));
    }

    bool on_step(const unsigned int t) const {
        return a.on_step(t) || b.on_step(t);
    }

    void at_start_step(const unsigned int t, const view_constgbx,
                      const view_constsupers supers) const {
        a.at_start_step(t, gbxs, supers);
        b.at_start_step(t, gbxs, supers);
    }

    Monitor auto get_monitor() const { return mo; }
};

auto operator>>(const Observer auto obs1, const Observer auto obs2) {
    const Monitor auto mol2 =
        CombinedMonitor{obs1.get_monitor(), obs2.get_monitor()};
    return CombinedObserver{obs1, obs2, mol2};
}

// ... //

/* use a combination of observers */
const Observer auto obs1 = StreamOutObserver(config);
const Observer auto obs2 = TimeObserver(config, dataset);
const Observer auto obs3 = StateObserver(config, dataset);
const Observer auto obs4 = SuperdropsObserver(config, dataset);
const Observer auto observer =
    obs4 >> obs3 >> obs2 >> obs1; // change this line to change observer
}

```

Listing 6: Example code showing how observers are combined in Cleo. Here we choose “ $\gg$ ” to be the operator in the definition of $A \oplus B = C$ for observers. Monitors are also monoid, with their own C++20 concept, associative operation, and identity element, but we omit their definition here for brevity.

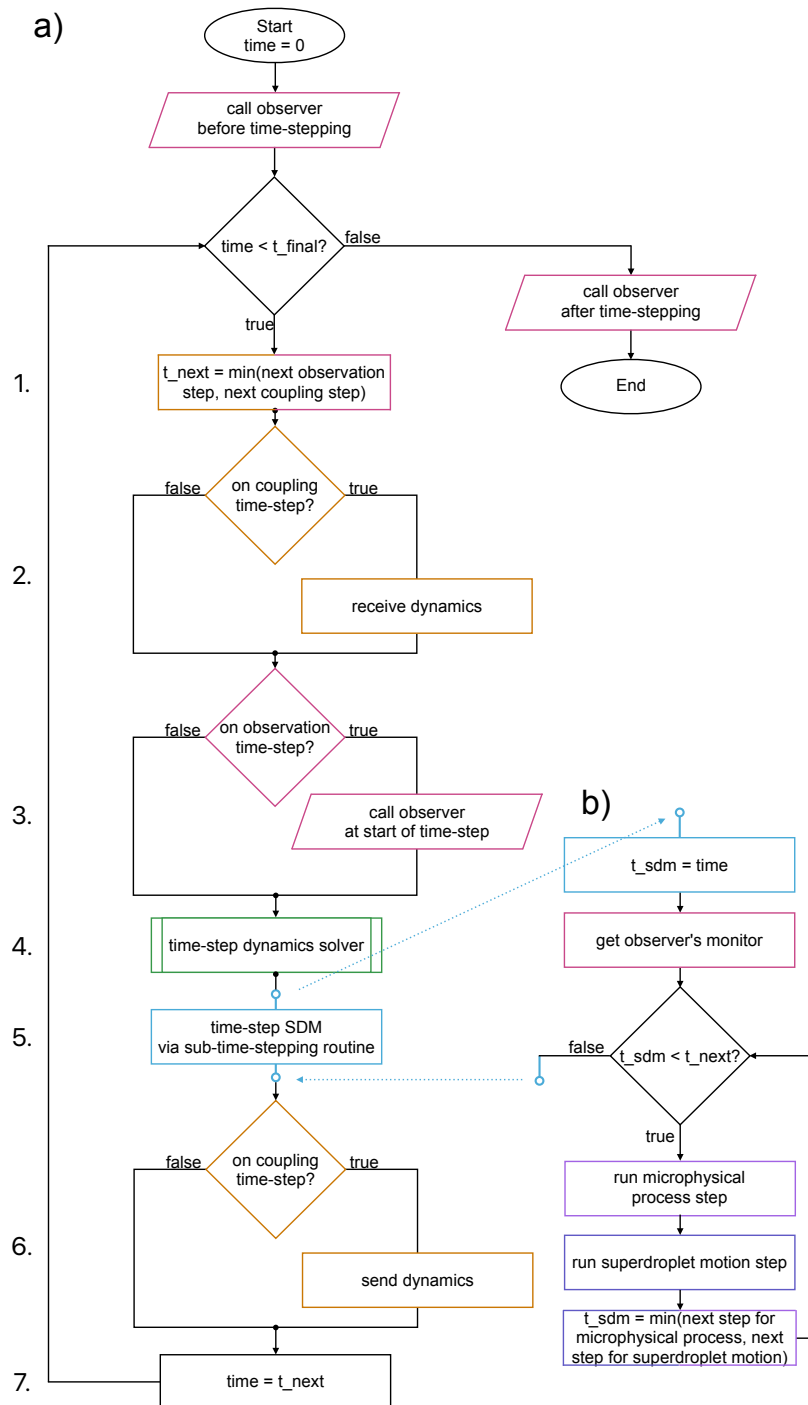


Figure A.9: (a) The flow diagram for Cleo’s time-stepping routine. Each step is coloured by which of Cleo’s structures it involves. The steps involved in the SDM sub-time-stepping routine are shown in (b) and are an extension of the logic presented in Figure A.4a to include output and superdroplet motion as well as microphysics.

	CPU-only	CUDA-enabled ²
CPU's	2 AMD 7763	2 AMD 7763
main memory / GB	256	512
number of threads (cores)	256 (128)	256 (128)
cores in (L3 cache, NUMA domain, single socket)	(8, 16, 64)	(8, 16, 64)
GPU's	-	1 NVIDIA A100
main memory / GB	-	40
compiler	intel-oneapi-compilers/2023.2.1-gcc-11.2.0	gcc/11.2.0-gcc-11.2.0
MPI	openmpi@4.1.5%oneapi/3cjsdq	openmpi@4.1.2%gcc@11.2.0
CMake	cmake@3.23.1%oneapi	cmake@3.26.3%gcc@=11.2.0/fuvwuhz
CUDA		cuda@12.2.0%gcc@=11.2.0

Table A.4: Configuration and software of Levante HPC for the performance tests of Cleo on a single node with only CPUs (for Serial, OpenMP, and C++Threads builds), and with CUDA enabled.

	Time-step / s	Additional Configuration Notes
droplet motion	3	terminal velocity parametrisation from Rogers et al. [217]
condensation/evaporation	1 (default)	minimum sub-time-step: 0.001 s
collision-coalescence	1	absolute and relative tolerances: 0.01 and 0.0 hydrodynamic kernel from Simmel et al. [245]

Table A.5: SDM configuration for the tests of Cleo’s superdroplet- and strong-scaling.

CLEO: THE NUMERICAL METHODS OF A NEW SUPERDROPLET MODEL INCLUDING A DROPLET BREAKUP ALGORITHM

The work in this appendix is currently under review at *Geoscientific Model Development* as C. J. A. Bayley et al., ‘CLEO: The Numerical Methods of a New Superdroplet Model including a Droplet Breakup Algorithm’, *EGU-sphere* **2025, 1–26 (2025).**

Authors: Clara J.A. Bayley^{1,2}, Ann Kristin Naumann^{1,3,4}, Florian Poydenot³,
Raphaella Vogel³, Bjorn Stevens¹, and Shin-Ichiro Shima⁵

¹Max-Planck-Institut für Meteorologie, Hamburg, Germany

²International Max Planck Research School on Earth System Modelling,
Hamburg, Germany

³Meteorologisches Institut, Universität Hamburg, Hamburg, Germany

⁴Ludwig-Maximilians-Universität München, Munich, Germany

⁵Graduate School of Information Science, University of Hyogo, Kobe, Ja-
pan

Author contributions: CJAB is the creator of Cleo and main developer, she also wrote and edited the manuscript. AKN and RV supervised the project, providing direction and support, as well as teaching CJAB about various aspects of cloud physics included in this paper. BS conceptualised the project, and had many discussions with CJAB which shaped the writing of the paper and provided scientific and technical support. He also helped analyse the test case results. FP contributed the parametrisation for the ventilation factor in condensation/evaporation. SIS supervised Cleo’s development, and contributed ideas and advice to the numerical methods for droplet breakup and condensation/evaporation. SIS also helped design the test cases for the numerical methods and analyse the results. AKN, FP, RV, BS and SIS all gave extensive feedback which contributed to the writing of the manuscript.

B.1 ABSTRACT

The numerical methods of conventional Eulerian models have obscured our fundamental understanding of cloud microphysics and introduced artificial uncertainties. In contrast, the Super-Droplet Model (SDM) provides a transparent link between model and theory, and remedies the numerical artifacts that hindered decades of cloud modelling. In light of its numerous advantages we've created a novel SDM for warm-cloud microphysics called Cleo, with the goal of making it feasible to run Large-Eddy Simulations (LES) with SDM in domains large enough to resolve shallow mesoscale cloud organisation $O(100\text{ km})$. Here we document the microphysics grounding Cleo and how it is translated into numerical methods, with the intention of assisting the physical interpretation of future LES and comparison with observations. We highlight subtle but important points where we differ from existing SDMs: in how we model the ventilation effect on evaporation, and how we account for uncertainty in our knowledge of droplet collisions. As well as modelling collision-coalescence we propose a low-cost extension to the original SDM algorithm which adds both collisional rebound and breakup. We demonstrate Cleo's capabilities with known test-cases for condensation/evaporation, collisions between droplets, and droplet motion, including an integrated test using the 1-D Kinematic Driver framework. Cleo can therefore now be used stand-alone, one-way coupled for piggybacking, or two-way coupled for LES, as a fully-functioning SDM capable of representing all the main microphysical processes driving warm-clouds.

B.2 INTRODUCTION

Substantial differences exist between conventional Eulerian models for cloud microphysics. Excessive degrees of freedom within one- and two-moment bulk schemes, the least expensive of existing microphysics models, allow for varying definitions of condensate categories, as well as the parametrisations for how they interact with one another, and their assumed size distributions [e.g. 130]. Increasing in sophistication, even the most conventional spectral bin schemes still vary in how they define condensate categories and their parametrisations. They also introduce differences in numerical implementations, for example to tackle numerical diffusion [182], such that even if two bin schemes solve the same equations they could calculate different results. These differences matter for more than just microphysics.

The differences between conventional Eulerian models have a wide-range of macrophysical ramifications. In global storm resolving models [e.g. 37, 176], only changing fall velocity parameters in one- and two-moment bulk schemes can substantially alter the relationship between moisture and convection, which has consequences for the Madden-Julian oscillation and the radiation balance of the entire atmosphere [185, 268, 272]. Varying definitions of condensate size distributions are also problematic because they alter shallow-cloud fraction and albedo, even in non-precipitating clouds, and their underlying assumptions are inconsistent with even the most basic kinematic processes [109, 110]. When Large-Eddy Simulations (LES) and observations are compared, differing precipitation patterns are frequently attributed to differences in microphysics schemes, and bin schemes are just as disparate as bulk ones [131, 232, 289]. Indeed there is extensive literature showing the many ways in which microphysics can appreciably effect

larger-scale dynamics [e.g. 18, 67, 90, 121], and thus as long as there are differences between microphysics schemes we can expect there to be differences in macrophysical outcomes. Differences inevitably arise from gaps in our knowledge of cloud microphysics, however, the differences highlighted above are not caused by knowledge gaps, but rather by “model uncertainty”, uncertainty inherent to *the way* that conventional Eulerian models represent what we already understand.

The Super-Droplet Model [SDM; 241] eradicates the biggest sources of conventional model uncertainty. SDM is a Lagrangian particle model for cloud microphysics where the condensate population is modelled by “superdroplets”. Superdroplets have the same properties as real condensates as well as an extra one, the multiplicity, which states how many real condensate particles that superdroplet represents. SDM therefore makes no assumptions about condensate categories or size distributions and it does not suffer from same intrinsic problems as bin schemes, such as numerical diffusion and the “curse of dimensionality” [84]. In the same way the accuracy of the condensate size distributions is bounded by the number of moments in a bulk scheme, and the number of bins in a bin scheme, it is the number of superdroplets which limit the accuracy of SDM. However SDM has an important convergence property. As the number of superdroplets increases, a simulation tends towards what we expect from explicit individual particle simulations, and the mean of many simulations approaches the solution to the Smoluchowski equation [247]. This is in stark contrast to bulk and bin schemes, where increasing refinement does not converge to a generic solution because the field-based description of particles has conceptual and numerical flaws [84, 182]. Moreover, the link between our fundamental understanding of microphysical processes and their representation in SDM is clear and direct, meaning that variations between SDMs hold a strong relation to differences in our understanding of the underlying physics.

We have created a novel computational implementation of SDM, Cleo, for a simplistic representation of warm-cloud microphysics well-suited to high-performance computers. In the companion to this paper we described the fundamental computational design of our SDM and its computational performance [22]. In this paper we document Cleo’s numerical methods for warm-cloud microphysics and demonstrate their behaviour in known test-cases. Since the link between the numerical methods and fundamental theory is so clear in SDM, by carefully describing our numerical methods we assist the physical interpretation of studies using Cleo and comparisons with other microphysics models.

Cleo models all the major microphysical processes of warm-clouds: condensation/evaporation, collisions between droplets and droplet motion. Most of the methods are already found in the literature, but in addition to documenting which of those Cleo adopts, we highlight particular novelties of our approaches and explain how the flexibility we incorporated reflects the uncertainties in our current understanding of warm-cloud microphysics. More specifically, we clarify the methods of Shima et al. [241] and Matsushima et al. [165] for modelling condensation/evaporation and extend them to include ventilation effects. For collision between droplets, Cleo facilitates various definitions for the collision kernel and the outcome of collisions between droplets. The breakup of raindrops as a consequence of collisions is included in De Jong et al. [51] and Bringi et al. [38], but is neglected by the majority of SDMs (and some bulk and bin microphysics schemes), des-

pite the fact it could be an important influence on the rate of evaporation in downdraughts [178, 180, 203, 235, 265]. Cleo therefore offers an extended version of the Shima et al. [241] collision-coalescence algorithm to permit the study of rebound and breakup.

In this paper the i^{th} superdroplet is defined by its spatial coordinates, $\vec{x}_i$, its multiplicity, ξ_i , and its attributes, $\vec{a}_i$, wherein R_i and $M_{s,i}$ are its radius and its mass of solute (aerosol), respectively. The total mass of each superdroplet is given by

$$M_{T,i} = M_{s,i} \left(1 - \frac{\rho_l}{\rho_{s,i}}\right) + \frac{4}{3} \pi \rho_l R_i^3, \quad (\text{B.1})$$

to account for the solute's volume whilst neglecting its change due to dissolution, where ρ_l and $\rho_{s,i}$ are the densities of liquid water and the solute, respectively. $\vec{a}_i$ symbolises $\vec{x}_i$ as well as all the superdroplet's intensive attributes, for example $\rho_{s,i}$.

Cleo's numerical methods for microphysics and superdroplet motion are found in Sections B.3 to B.5, and their validations are found in Section B.6. Section B.3 describes how Cleo models condensation/evaporation, including ventilation effects and an adaptive sub-time-stepping algorithm. Section B.4 explains how we have extended the collision-coalescence algorithm of Shima et al. [241] to include a framework for collisional breakup and rebound, and Section B.5 explains how Cleo determines superdroplet motion.

B.3 CONDENSATION/EVAPORATION

We use Köhler theory to describe droplet growth due to evaporation/condensation, but also account for the enhancement of evaporation of moving droplets by including a ventilation factor. The change of the i^{th} superdroplet's radius, R_i , is therefore

$$R_i \frac{dR_i}{dt} = f_v \frac{(S - 1) - \frac{a}{R_i} + \frac{b}{R_i^3}}{F_k + F_d}, \quad (\text{B.2})$$

where f_v is the ventilation factor; S is the ambient saturation ratio of the grid-box; F_k and F_d are, respectively, the heat conductivity and vapour diffusion factors; $\frac{a}{R_i}$ represents the effect of curvature on the saturation at a droplet's surface; and $\frac{b}{R_i^3}$ accounts for the reduction in water vapour pressure due to the presence of solute. We use the same formulations for F_k , F_d , a and b as in Shima et al. [241] [134, 218]. Supersaturation fluctuations caused by sub-grid-scale turbulence are not taken into account.

We obtain the ventilation factor, f_v , by fitting the curve

$$f_v(R) = 1 + \left(\frac{1}{\alpha_1 R^{\beta_1}} + \frac{1}{\alpha_2 R^{\beta_2}} \right)^{-1} \quad (\text{B.3})$$

to the curve from Pruppacher and Klett [208], such that $\alpha_1 = 6.954 \times 10^7 \text{ m}^{-\beta_1}$, $\alpha_2 = 1.069 \times 10^3 \text{ m}^{-\beta_2}$, $\beta_1 = 1.963$, $\beta_2 = 0.702$. This gives similar values as Pruppacher and Klett [208] [given the terminal velocity parametrisation from 208] but makes the size dependence more explicit whilst neglecting the small dependence of f_v on pressure variations. Additionally we constrain $f_v \leq 20.0$ to reflect the fact that droplets with a radius larger than

approximately 3 mm have the same terminal velocity. As shown in Figure B.1, f_v is negligible for droplets with radii less than about 0.1 mm, but can increase the evaporation rate of large droplets by an order of magnitude.

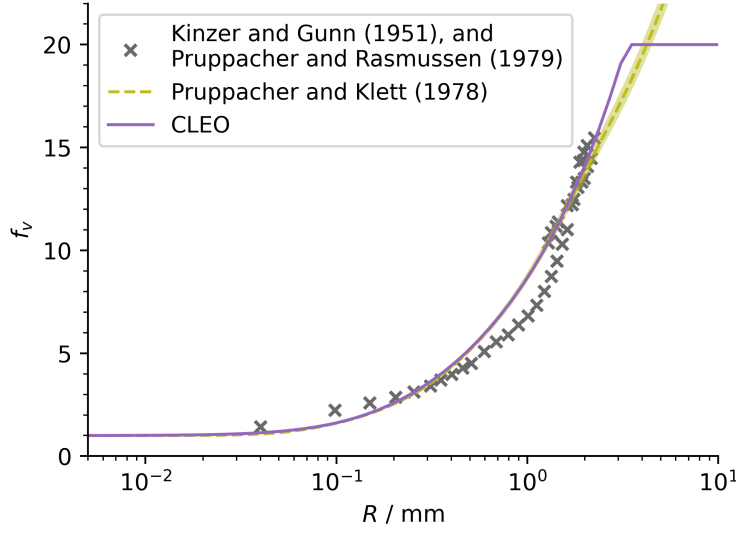


Figure B.1: The ventilation factor we use in Cleo compared to the experimental data from Kinzer and Gunn [132], and Pruppacher and Rasmussen [209], and compared to the fit to the data from Pruppacher and Klett [208] at $T = (288.15 \pm 15.00)$ K in a standard atmosphere.

Careful consideration of the numerical methods for integrating Equation (B.2) is required because this ordinary differential equation is stiff. We use an implicit Euler method with a Newton-Raphson root-finding algorithm similar to Shima et al. [241] to ensure stability when integrating the modified version of the ODE in Equation (B.2). Our method saves computational cost by permitting a comparatively large default time-step ($\Delta t_{\text{cond}} \sim 1$ s) and very often keeping the number of Newton-Raphson iterations below three. Explicitly, given the time-step for condensation/evaporation, Δt_{cond} , and the notation $R_i(t^n) = R_i^n$, we make the approximation $f_v(R) \approx f_v(R_i^n)$ and let $f_v(R_i^n) = f_v$, the ODE is discretised as

$$0 = \frac{Z - (R_i^n)^2}{\Delta t_{\text{cond}}} - \frac{2f_v[(S-1) - aZ^{-1/2} + bZ^{-3/2}]}{F_k + F_d}; \quad (\text{B.4})$$

where $Z = (R_i^{n+1})^2$. The Newton-Raphson method is then used to find the real positive root of the polynomial,

$$g(Z) = \frac{Z - (R_i^n)^2}{\Delta t_{\text{cond}}} - \frac{2f_v[(S-1) - aZ^{-1/2} + bZ^{-3/2}]}{F_k + F_d}; \quad (\text{B.5})$$

such that the radius of each superdroplet at the subsequent time-step is given by $R_i^{n+1} = +\sqrt{Z}$ when $g(Z) = 0$. We use $g(Z)$ rather than the higher order polynomial in terms of $+\sqrt{Z}$ because it converges more rapidly — in fact in the limit $a = b = 0$, the Newton-Raphson method can find the true root of Equation (B.5) with one iteration. A reasonable initial guess close to the true solution for $g(Z) = 0$ can make convergence fast and so we choose

our first guess for Z to be $(R_i^n)^2$ unless $S > 1 + \sqrt{\frac{4a^3}{27b}}$, in which case the superdroplet is already activated and we make the initial guess very large: at least 10^{-6} m^2 .

There are upto three real roots of $g(Z)$ and so we must ensure that the genuine solution for R_i^{n+1} is converged upon. If the droplet was previously and is currently un-activated, and in an environment with a supersaturation less than its activation supersaturation, i.e. if

$$S \leq 1 + \sqrt{\frac{4a^3}{27b}}, \quad \text{and} \quad R_i^n < \sqrt{\frac{3b}{a}}, \quad (\text{B.6})$$

then the uniqueness of the solution to $g(Z) = 0$ is guaranteed. Likewise, if the time-step is small enough,

$$\Delta t_{\text{cond}} \leq \frac{25b(F_k + F_d)}{2a^2 f_v} \sqrt{\frac{5b}{a}}, \quad (\text{B.7})$$

uniqueness is guaranteed (as in Matsushima et al. [165] modulo the ventilation factor). In either of these cases we attempt two iterations of the Newton-Raphson method and then perform a standard local error test for convergence, whereby the method has converged if $g(Z) < \alpha + \beta g(R_i^n)^2$ for some absolute and relative tolerances, α and β , respectively¹. R_i^{n+1} has been found if the test passes, whereas if it fails we perform further iterations and check for convergence after each one. However, to prevent parasitic cases from running for infinite time, the simulation is terminated if a chosen maximum number of iterations is exceeded¹. When neither condition for a unique solution is met adaptive time-stepping is performed, whereby Δt_{cond} is divided into sub-steps which each obey the second criteria for uniqueness. It is highly improbable but these sub-steps can be extremely small [165]. We therefore allow a minimum sub-step to be set in order to reduce the cost of the simulation but at the risk of finding the incorrect solution for R_i^{n+1} .

B.4 COLLISIONS

Modelling collisions between droplets in Cleo can be done with or without collisional breakup and rebound. Collisional breakup during rainfall has been observed in situ and has been suggested as an important control on the timing and intensity of rain in LES [237, 275]. Since breakup converts large droplets into significantly smaller ones, its effect on the higher moments of the droplet size distribution can be appreciable [e.g. 167] [references in 182]. This motivates two alternative treatments for collisions in Cleo, schematised in Figure B.2, which are:

- (a) the original algorithm of Shima et al. [241] which models only collision-coalescence ,
- (b) the extended algorithm which also includes rebound and breakup .

In both treatments, the steps to determine whether or not a collision occurs is the same as in Shima et al. [241]. All the superdroplets in a grid-box are first randomly paired with one-another. Then for each pair the probability that they collide is compared to a random number, ϕ_α , to determine

¹ By default we choose $\alpha = 0.01$ and $\beta = 0.0$, and the maximum number of iterations as 50.

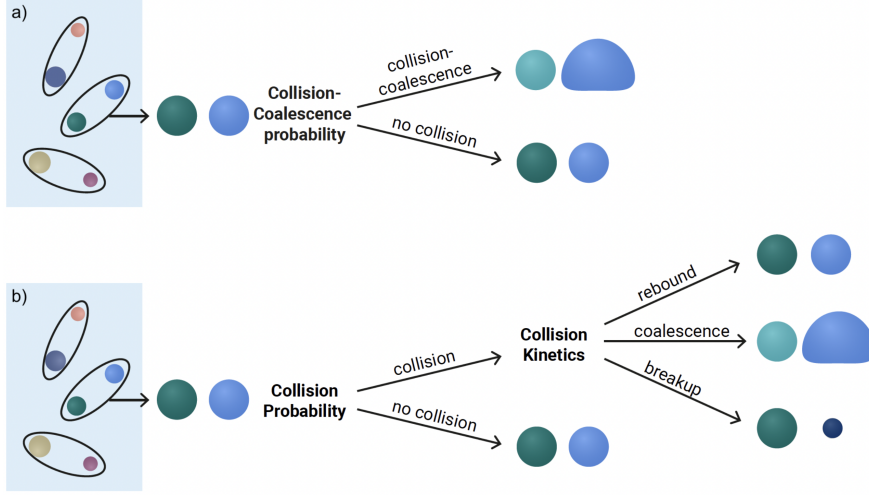


Figure B.2: (a) Schematic of the original collision-coalescence algorithm from Shima et al. [241]. (b) The extended collision algorithm including breakup and rebound as well as coalescence. A lightening/darkening of a superdroplet indicates a decrease/increase in its multiplicity.

whether or not a collision is enacted. The probability that two superdroplets collide, p_α , is the probability that two real droplets collide, P_{jk} , scaled by a factor dependent on the superdroplets' multiplicities. Usually P_{jk} can be calculated given a collision kernel, K_{jk} , as

$$P_{jk} = K_{jk} \frac{\Delta t_{\text{coll}}}{\Delta V}, \quad (\text{B.8})$$

for a pair of droplets j and k , where Δt_{coll} and ΔV are the collision time interval and volume, respectively [241]. The probability two superdroplets collide is then

$$p_\alpha = \max(\xi_j, \xi_k) \frac{n_s(n_s - 1)/2}{[n_s/2]} P_{jk}, \quad (\text{B.9})$$

where n_s is the number of superdroplets in the collision volume, ΔV .

Cleo defines the collision probability flexibly to make it easy to switch between different formulations of P_{jk} . There are large uncertainties in determining the collision probability of a pair of superdroplets' because the collision kernel is so poorly constrained. As such, several formulations for K_{jk} exist, which differ for example in their treatment of turbulence [e.g. 15, 91, 159]. The calculation for P_{jk} in Cleo can therefore be changed before compilation as long as it obeys the C++20 concept, [112] [see also 22], we prescribe for the collision probability. There are two P_{jk} calculations implemented at the time of writing: one for Golovin's kernel [75], and the other for the hydrodynamic kernel with the formulation for the terminal velocity from Simmel et al. [245] and collision efficiencies from Long [159], as also done by Grabowski and Wang [76].

If a collision occurs, treatments (a) and (b) differ in their outcome. In the original algorithm, coalescence is assumed and the superdroplets are updated as in Section B.4.1. Whereas if a collision occurs in the extended algorithm, the last step of the original collision-coalescence algorithm is modi-

fied as shown in Figure B.2b. Rather than assuming a successful collision of two superdroplets results in coalescence, the attributes of the original superdroplets are used to determine if the collision causes rebound, coalescence, or breakup. We then modify the superdroplets to reflect the appropriate outcome: as in Section B.4.1 for coalescence and as in Section B.4.2 for breakup. In the event of rebound, the superdroplets remain unchanged.

Cleo has several options for how the attributes of colliding superdroplets determine whether the outcome of the extended algorithm is rebound, coalescence, or breakup. This is physically motivated since properties of droplets, such as their velocities, shapes, and masses, determine whether a collision has sufficient energy to result in coalescence or breakup instead of rebound, but exactly how such properties determine the outcome of a collision is disputed. Our algorithm can therefore easily interchange various formulations from the literature, at the time of writing from Low and List [161], Testik et al. [274], and Szakáll and Urbich [270].

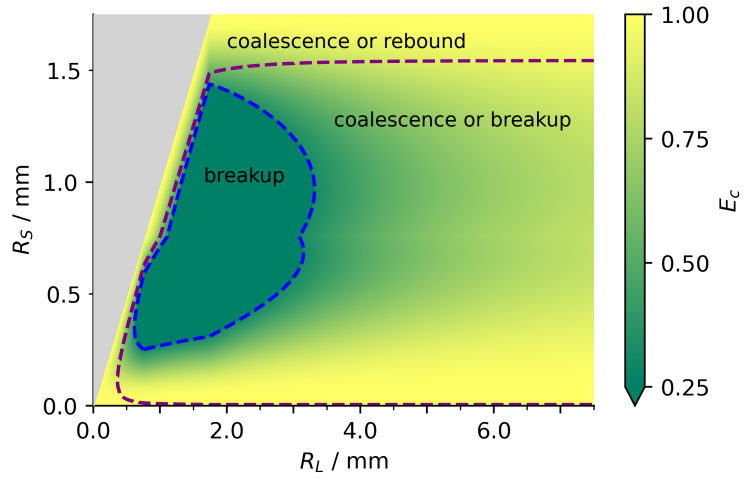


Figure B.3: The probability, based on Testik et al. [274], for whether a collision between two droplets results in coalescence, rebound or breakup. Inside the blue dashed contour it is certain that a collision results in breakup, between the blue and purple contours it may result in coalescence or breakup, and outside the purple contour it may result in coalescence or rebound. The colour gives the coalescence efficiency, in other words the probability of coalescence given that a collision occurs, according to Straub et al. [266].

As an example, Figure B.3 demonstrates how the outcome of a collision is determined based on Testik et al. [274]. There are three possible regimes depending on the magnitude of the collision kinetic energy, T_E , relative to

the surface tension energy of the smaller and larger droplet, S_S and S_L , respectively [276]. Using the definitions

$$T_E = \frac{2}{3}\pi\rho_l \frac{R_j^3 R_k^3}{R_j^3 + R_k^3} |\vec{v}_{j,\infty} - \vec{v}_{k,\infty}|^2, \quad (\text{B.10})$$

$$S_i = 4\pi\sigma_l R_i^2, \quad (\text{B.11})$$

$$R_S = \min(R_j, R_k), \quad (\text{B.12})$$

$$R_L = \max(R_j, R_k), \quad (\text{B.13})$$

where ρ_l is the density and σ_l is the surface tension of liquid water, these three regimes are:

1. when $T_E < S_S$ either coalescence or rebound occurs,
2. when $S_S < T_E < S_L$ either coalescence or breakup occurs,
3. when $T_E > S_L$ breakup occurs.

In regimes 1 and 2 a coalescence efficiency, $E_c \in [0, 1]$, gives the probability two droplets coalesce, $P_{c,jk}$, given that they collided, i.e. $P_{c,jk} = E_c P_{jk}$. Here, the parametrisation for E_c is

$$E_c = e^{-1.15W}, \quad (\text{B.14})$$

where the Weber number, $W = \frac{T_E}{S_c}$, and $S_c = 4\pi\sigma(R_S^3 + R_L^3)^{2/3}$; as argued for by Straub et al. [266]. Unlike De Jong et al. [51], we rescale the original random number which was used to determine that a collision occurred rather than draw a new random number, $\phi'_\alpha = \phi_\alpha / ([p_\alpha] - p_\alpha)$, and then we compare ϕ'_α with E_c to decide if the collision results in coalescence or the alternative.

B.4.1 Coalescence

The change in attributes of a pair of superdroplets that undergo coalescence follows Shima et al. [241] and is illustrated in Figure B.4.

Given a pair of superdroplets (j, k) we can choose without loss of generality the multiplicity of the j^{th} superdroplet to be at least as large as that of the k^{th} , i.e. $\xi_j \geq \xi_k$. The permitted number of consecutive coalescence events, $\tilde{\gamma}_\alpha$, is then used to determine how the superdroplet attributes are changed to enact coalescence. To calculate $\tilde{\gamma}_\alpha$, we compare p_α to the random number ϕ_α as well as the multiplicities of the superdroplets since

$$\gamma_\alpha = \begin{cases} [p_\alpha] + 1 & \text{if } \phi_\alpha < p_\alpha - [p_\alpha], \\ [p_\alpha] & \text{if } \phi_\alpha \geq p_\alpha - [p_\alpha]; \end{cases} \quad (\text{B.15})$$

$$\tilde{\gamma}_\alpha = \min(\gamma_\alpha, \lfloor \xi_j / \xi_k \rfloor). \quad (\text{B.16})$$

There are then two possible scenarios which enact coalescence:

- (a) if $\xi_j > \tilde{\gamma}_\alpha \xi_k$, the less multiplicitous superdroplet grows by consuming $\tilde{\gamma}_\alpha \xi_k$ droplets from the other superdroplet,

$$\xi'_j = \xi_j - \tilde{\gamma}_\alpha \xi_k, \quad \xi'_k = \xi_k, \quad (\text{B.17})$$

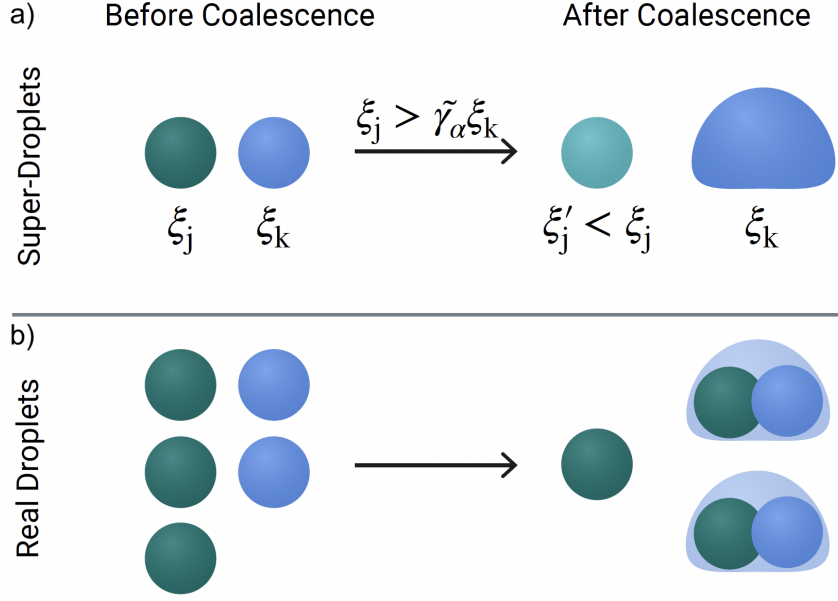


Figure B.4: Coalescence as in Shima et al. [241]. (a) The superdroplet representation of coalescence whereby the multiplicity of the more multiplicitous droplet decreases to increase the size of the other superdroplet. A lightening/darkening of a superdroplet indicates a decrease/increase in its multiplicity. (b) The real droplet equivalent of the superdroplets above when $\xi_j = 3$, $\xi_k = 2$ and $\tilde{\gamma}_\alpha = 1$.

$$R'_j = R_j, R'_k = (\tilde{\gamma}_\alpha R_j^3 + R_k^3)^{1/3}, \quad (\text{B.18})$$

$$M'_{s,j} = M_{s,j}, M'_{s,k} = (\tilde{\gamma}_\alpha M_{s,j} + M_{s,k}), \quad (\text{B.19})$$

$$\vec{a}'_j = \vec{a}_j, \vec{a}'_k = \vec{a}_k; \quad (\text{B.20})$$

(b) otherwise $\xi_j = \tilde{\gamma}_\alpha \xi_k$, and rather than ending up with $\xi'_j = 0$ both superdroplets are used to represent coalesced droplets,

$$\xi'_j = \lfloor \xi_k/2 \rfloor, \xi'_k = \xi_k - \xi'_j, \quad (\text{B.21})$$

$$R'_j = R'_k = (\tilde{\gamma}_\alpha R_j^3 + R_k^3)^{1/3}, \quad (\text{B.22})$$

$$M'_{s,j} = M'_{s,k} = (\tilde{\gamma}_\alpha M_{s,j} + M_{s,k}), \quad (\text{B.23})$$

$$\vec{a}_j' = \vec{a}_j, \vec{a}_k' = \vec{a}_k; \quad (\text{B.24})$$

In the exceptional case that $\xi_j' = 0$, we currently choose to terminate the simulation rather than remove the superdroplet.

B.4.2 Breakup

To keep the computational advantages of the original SDM collision algorithm, the fragments created from one breakup event are represented by a single superdroplet. In reality, collisional breakup results in a spectrum of different fragments whose likelihood depends on the two droplets that collided. However the number of simulated particles must be unchanged during collisions to ensure SDM is manageable, and thus the fragments from breakup must all have the same attributes. Ideally these would be randomly sampled from a known fragment distribution, as in De Jong et al. [51] and Brongi et al. [38]. However observations of these distributions are limited by both resolution and the undersampling of initial droplet sizes [161, 162, 205, 230, 270].

Given the acute shortage of empirical data and our desire for high computational performance, our algorithm does not randomly sample a known fragment distribution. Instead the properties describing the fragments from collisional breakup, ξ_{frag} , R_{frag} and $M_{\text{s,frag}}$, are determined by the initial superdroplet attributes. We provide several ways for how to do this in Cleo; for example, one way asserts mass conservation and that the number of fragments depends on a collision's kinetic energy based on Schlottke et al. [230], as detailed in Appendix B.10.1. Such a method does not seek to reproduce the measured fragment distributions, but rather pursue an overtly simple approach with the intention of exploring whether and how breakup could impact observables such as rain evaporation rates and radar reflectivity signals [e.g. 40, 182].

Two previous papers have introduced a collisional breakup algorithm into SDM [38, 51]. Unlike the breakup algorithm in McSnow [38], we ensure the conservation of superdroplet number and thus computational tractability. Our algorithm is similar to that of pySDM [51], but with two adjustments. Firstly, whether a collision results in rebound, coalescence, or breakup is deterministic in our algorithm, whereas an additional probabilistic step is used in De Jong et al. [51]. Secondly, in the event of breakup we avoid a recursive algorithm by prohibiting multiple breakup events in a single time-step. In the original algorithm of Shima et al. [241], consecutive coalescence events can be enacted in a single time-step via the factor for the permitted number of consecutive coalescence events, $\tilde{\gamma}_\alpha$. Allowing consecutive coalescence rests on the assumption that after the first coalescence event(s) the droplets are similar enough that the coalescence probability is approximately unchanged, meaning the collision probability and its outcome do not need recalculation before subsequent events. In our breakup algorithm we make the opposite assumption: that after a single breakup event there is negligible probability of the resultant droplets undergoing further collisional breakup with one-another and thus only one breakup event can occur per time-step ($\tilde{\gamma}_\alpha = 1$). This is a reasonable assumption for the $O(1\text{ s})$ time-step required

for collision-coalescence and is a necessary reduction to the computational cost of the algorithm in large simulations.

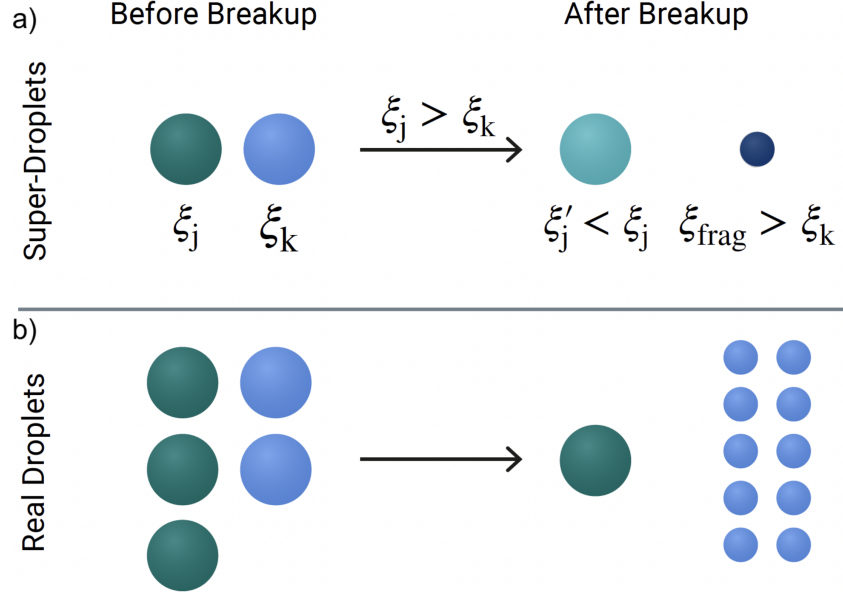


Figure B.5: (a) The superdroplet representation of breakup decreases the multiplicity of the more multiplicitous droplet to fragment the other droplet. A lightening/darkening of a superdroplet indicates a decrease/increase in its multiplicity. (b) The real droplet equivalent of the superdroplets above when $\xi_j = 3$, $\xi_k = 2$, and $\xi_{frag} = 10$.

The change in superdroplets that undergo collisional breakup is illustrated in Figure B.5 and has two cases analogous to coalescence:

- (a) if $\xi_j \neq \xi_k$, we can choose without loss of generality $\xi_j > \xi_k$ such that ξ_k droplets from the more multiplicitous superdroplet are involved in the breakup with the other superdroplet,

$$\xi'_j = \xi_j - \xi_k, \quad \xi'_k = \xi_{frag}, \quad (B.25)$$

$$R'_j = R_j, \quad R'_k = R_{frag}, \quad (B.26)$$

$$M'_{s,j} = M_{s,j}, \quad M'_{s,k} = M_{s,frag}, \quad (B.27)$$

$$\vec{a}'_j = \vec{a}_j, \quad \vec{a}'_k = \vec{a}_k; \quad (B.28)$$

- (b) if $\xi_j = \xi_k$, both superdroplets are used to represent the result of breakup,

$$\xi'_j = \text{round}(\xi_{frag}/2), \quad \xi'_k = \xi_k - \text{round}(\xi_{frag}/2), \quad (B.29)$$

$$R'_j = R'_k = R_{\text{frag}}, \quad (\text{B.30})$$

$$M'_{s,j} = M'_{s,k} = M_{s,\text{frag}}, \quad (\text{B.31})$$

$$\vec{a}'_j = \vec{a}_j, \quad \vec{a}'_k = \vec{a}_k; \quad (\text{B.32})$$

B.5 MOTION

We use a simple predictor-corrector method to model superdroplet motion. One of the most computationally expensive parts to Lagrangian microphysics is the motion of particles throughout the domain [22, 165]. However, since a superdroplet's location is not the true location of the droplets it represents, there is no imperative to precisely resolve a superdroplet's position. Indeed looking at the conceptual picture of SDM, perhaps the most suitable choice for superdroplet motion would be probabilistic, similar to Curtis et al. [50]. However, we save expense by choosing Heun's method, a second-order predictor-corrector method, with a simple linear interpolation of the wind velocity. This method by construction preserves the divergence of the flow field, and gives physically consistent results at a 1 s time-step [78].

In summary, we integrate the equation for the coordinates $\vec{x}_i$ of the i^{th} superdroplet,

$$\frac{d\vec{x}_i(t)}{dt} = \vec{u}_i(\vec{x}_i, t), \quad (\text{B.33})$$

where the superdroplet's velocity, $\vec{u}_i(\vec{x}_i, t)$, is

$$\vec{u}_i(\vec{x}_i, t) = \vec{w}(\vec{x}_i, t) + \vec{v}_{i,\infty}(\vec{x}_i), \quad (\text{B.34})$$

and $\vec{v}_{i,\infty}(\vec{x}_i) = -v_{i,\infty}(\vec{x}_i)\hat{z}$ is the superdroplet's terminal velocity, and $\vec{w}(\vec{x}_i, t)$ is the wind velocity, obtained by simple linear interpolation of the wind velocity at a grid-box's faces. In Cleo the terminal velocity is defined by a C++20 concept so that different formulations can be easily interchanged [112] [see also 22]. The options currently available are depicted in Figure B.6. Using the notation $\vec{x}_i(t^n) = \vec{x}_i^n$, and the time-step for superdroplet motion, Δt_m , the predicted coordinates at the subsequent time-step are

$$\vec{x}_i^{n+1} = \vec{x}_i^n + \vec{u}_i^n(\vec{x}_i^n) \Delta t_m, \quad (\text{B.35})$$

which are corrected to

$$\vec{x}_i^{n+1} = \vec{x}_i^n + \left[\vec{u}_i^n(\vec{x}_i^{n+1}) + \vec{u}_i^n(\vec{x}_i^n) \right] \frac{\Delta t_m}{2}. \quad (\text{B.36})$$

This formulation for particle motion could easily be extended to account for the effects of particle inertia and sub-grid scale turbulence, for example following Naumann and Seifert [187], which would increase the fidelity of the particle trajectories, especially for large droplets whose inertial relaxation timescale is larger than Kolmogorov time scale.

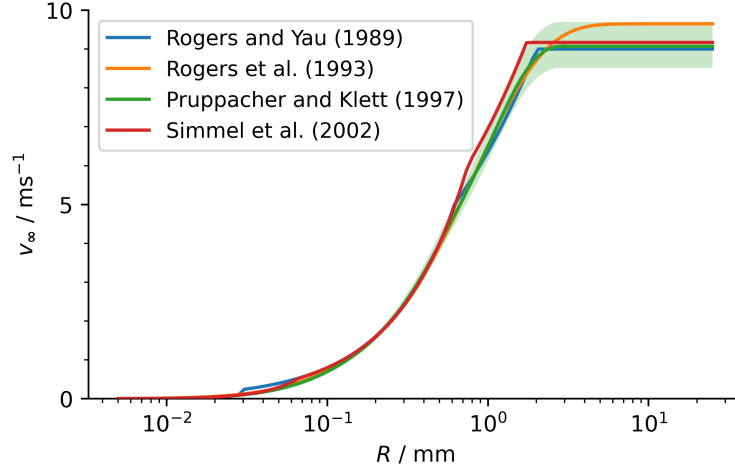


Figure B.6: The formulations for terminal velocity currently available in Cleo, and for comparison the formulation according to Pruppacher and Klett [208] at $T = (288.15 \pm 15.00)$ K in a standard atmosphere .

B.6 VALIDATIONS

This section presents three test-cases which demonstrate the behaviour of our numerical methods for condensation/evaporation, collisions between droplets, and droplet motion individually. Furthermore, we present a final integrated test-case applying all three methods at once using the 1-D Kinematic Driver (KiD) modelling framework [243, 244].

Figure B.9 shows the results from Cleo for the same test-case as described in Section 7 of Arabas and Shima [13] for the adiabatic expansion/contraction of a rising/falling volume of air through a hydrostatically equilibrated atmosphere. We used 0.0 and 0.01 for the relative and absolute tolerances of our ODE solver, and 50 and 1 ms for the maximum number of Newton-Raphson iterations and minimum sub-time-step, respectively. Each column shows the evolution of the air mass given a certain number concentration and/or initial radius of its dry aerosol mono-size distribution. In particular, it shows how the displacement, supersaturation, and droplet radii evolve given three different mean vertical velocities, $\langle w \rangle$, for the ascent/descent of the air mass. The first two columns show that, as expected, the faster the vertical velocity, the higher the dis-equilibrium supersaturation during ascent (and the lower the dis-equilibrium supersaturation during descent). By comparing the first two columns, we can also see that increasing the aerosol concentration lowers the dis-equilibrium supersaturation. All these trends in the supersaturation can be explained by faster vertical velocities and/or fewer aerosol causing the rate at which the relative humidity increases with height to be greater than the rate at which the droplets undergo condensation. Conversely, the lower the vertical velocity and higher the droplet number concentration, the closer the air mass to its equilibrium behaviour, where supersaturation does not exceed 0% and the radii follow the Köhler curve. In the third column however, the Köhler curve is not approached, because the radii and number concentration are low enough that cusp bifurcation occurs and a sudden “jump” in the droplet growth is observed instead. Due to differing numerics and initial thermodynamic conditions compared to Ara-

bas and Shima [13], the plots are not completely identical (we use 1000 hPa, 298.15 K, and 98.0%, for pressure, temperature and relative humidity, respectively), but our supersaturations are within 0.1% of theirs and we also observe cusp bifurcation, albeit at a slightly smaller radius of $0.03 \mu\text{m}$. Notwithstanding these differences, we can say that Cleo is able to reproduce the expected evolution of microphysical and thermodynamic conditions for all three mean vertical velocities.

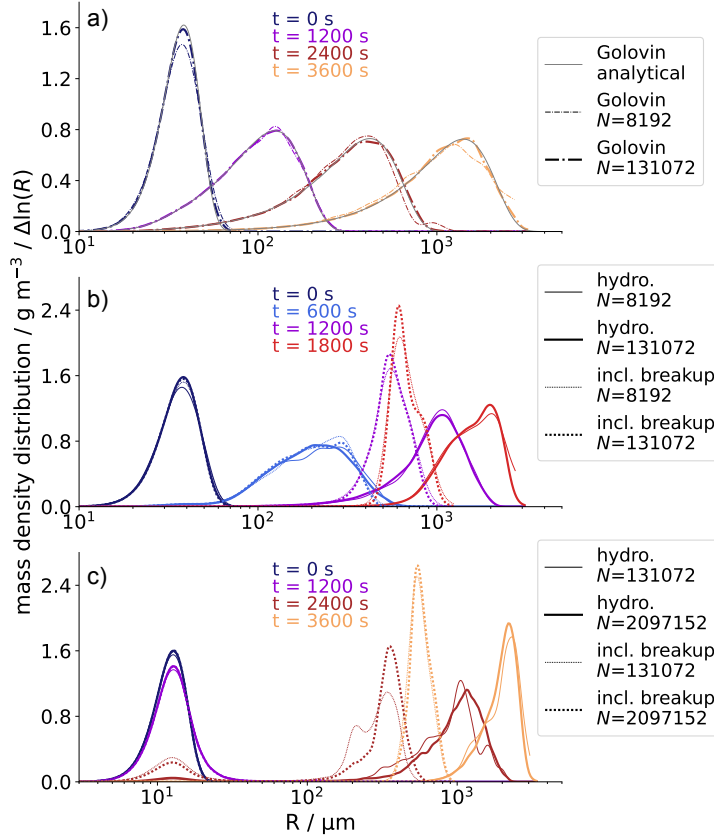


Figure B.7: Test of SDM collisions as in Shima et al. [241, Sec. 5.1.4, Figure 2]. The line-style indicates the collision kernel; the line-width indicates the number of superdroplets, N ; the colour indicates the time. Sub-figure (a) additionally shows the analytical solution for the Golovin kernel at each time (solid grey). The sub-figures are the same as in Shima et al. [241] except in panels (b) and (c), where, as well as the hydrodynamic kernel, we additionally plot the distribution's evolution when breakup and rebound are accounted for, based on the parametrisations from Schlottke et al. [230], Straub et al. [266], and Testik et al. [274] (see main text, Section B.4).

Figure B.7 shows the results from Cleo for the same test-case as described in Section 5.1.4 of Shima et al. [241] for the stochastic collisions between water droplets in an arbitrary volume of air. We show results for the same setups as Shima et al. [241], namely for the evolution of the mass density distribution (normalised by log-space bin-width) given different collision kernels and droplet populations whose volumes are initially exponentially distributed. Figure B.7a shows the evolution when the kernel from Golovin

[75] is used, whereas Figures B.7b-c show the evolution when the hydrodynamic kernel is used. For all cases, we reproduce the results of Shima et al. [241] when collisions are assumed to result in coalescence. We also demonstrate in Figures B.7b-c how the distribution evolves according to the extended framework from Section B.4, which also allows for collisional breakup and rebound. The probability of breakup and rebound was calculated from Testik et al. [274] and Straub et al. [266] as described in Section B.4, and if breakup occurred the fragment properties were determined based on the parametrisation of Schlottke et al. [230] (as described in Appendix B.10.1). Up to about 600 s the evolution of the droplet size distribution is comparable to the coalescence-only case, but the subsequent growth of droplets is much delayed and converges to a stationary distribution centred at $750\text{ }\mu\text{m}$, rather than $2400\text{ }\mu\text{m}$ in the coalescence-only case. As expected, collisional breakup slows and inhibits the growth of mm-size droplets. It also has a much larger impact on the distribution's evolution than changing the collision efficiencies from Long [159] to Hall [91] in the coalescence-only case. It remains to be seen how a kernel accounting for turbulent effects would behave, but that too would likely show considerably larger differences [76].

Figure B.10 shows the results from Cleo for the motion of particles in the 2-D kinematic laminar flow model described in Section 2.1 of Arabas et al. [11]. 128 superdroplets are initially randomly distributed throughout each grid-box in the domain and the flow is simulated for 1 hour with a time-step of 1 s for both motion and data output, and for three simulations with grid-spacing $\Delta x = \Delta z = 100\text{ m}$, 50 m , and 25 m , respectively. We do not simulate microphysics and the droplets have no fall velocity, so the superdroplets behave like tracers for the flow. As expected, Cleo's particle motion does indeed follow the flow field and respect its divergence-free properties, as confirmed by the conservation of particle number in each grid-box throughout the simulations (not shown). Figure B.10 also verifies that increasing the grid resolution increases the accuracy of the particle motion because the interpolation distances in the predictor-corrector algorithm reduce.

Figure B.8 demonstrates the integrated test of Cleo's numerical methods using the 1-D KiD framework. Here we setup the column as described by Shipway and Hill [243], with a 25 m grid-spacing and sinusoidal updraught velocity for the first 10 min with the maximum speed constant, $w_1 = 3\text{ m s}^{-1}$. We couple Cleo via python bindings to PyMPDATA [20] to advect water vapour, whereas we advect liquid water (superdroplets) within Cleo given the wind field at each time-step (1.25 s). We show results for a 10 member ensemble where for each member we initialise 50 cm^{-3} of dry aerosol throughout the domain by assigning 256 superdroplets per grid-box each a constant multiplicity and randomly sampling a single-mode lognormal distribution with a geometric mean diameter of $0.08\text{ }\mu\text{m}$ and a log standard deviation of 1.4, as chosen by Hill et al. [94]. Unlike both Shipway and Hill [243] and Hill et al. [94], we specify the initial air density profile to be consistent with hydrostatic equilibrium including the initial water vapour, with the surface pressure at 1000 hPa . The different microphysics schemes and density profiles causes slight differences between our clouds and those of Shipway and Hill [243], Hill et al. [94] and De Jong et al. [51], but nevertheless the qualitative behaviour of the test-cases are in agreement and our results lie within the spread amongst different SDMs in Hill et al. [94]. Figure B.8(a-b) shows the liquid water mass mixing ratio, q_l , when only condensation/evaporation is modelled. The cloud reaches a steady state im-

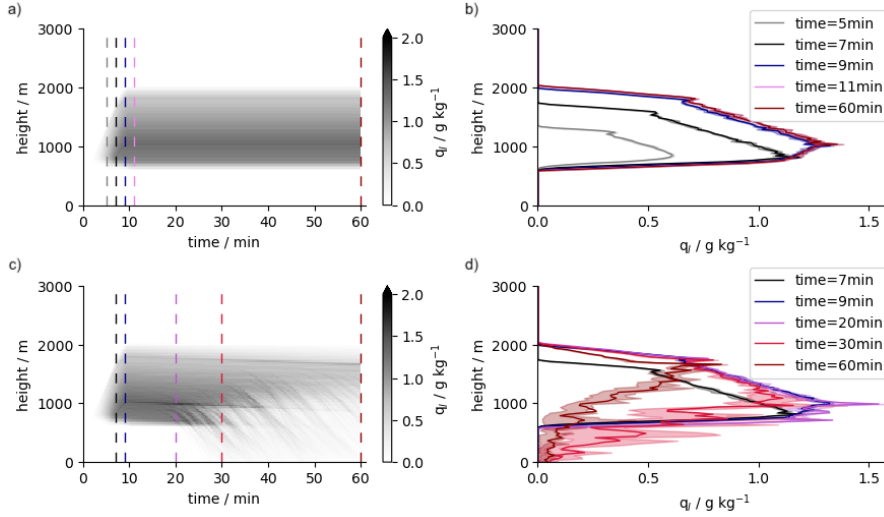


Figure B.8: Evolution of the liquid water mass mixing ratio, q_l , during the 1-D KiD test-case: (a-b) only modelling condensation/evaporation, (c-d) modelling condensation/evaporation, collision-coalescence and droplet sedimentation. Results are shown for the mean and interquartile range over an ensemble of 10 members, and figures b) and d) are cross sections across the vertical coloured lines in a) and c) respectively.

mediately after the updraught velocity terminates, with liquid water content peaking in the middle of our cloud, albeit at a higher value and lower height. Without collision-coalescence, there is no spread across the ensemble members, which indicates little sensitivity of condensation/evaporation to the sampling of the initial dry aerosol distribution, as reported by Hill et al. [94]. In contrast, Figure B.8(c-d) show the same setup when not only condensation/evaporation, but also collision-coalescence and droplet motion (including their fall velocities) are modelled, and in these simulations the ensemble spread is significant because of the stochasticity in the collision-coalescence algorithm. The mean precipitation onset is around 30 min, in agreement with the range of SDMs in Hill et al. [94], and likewise the evolution of the liquid water path, cloud droplet number concentration, mean volume diameter and its standard deviation also lie within the SDMs' spread (not shown). More exact agreement is hard to achieve given the ambiguity in the test-case setup and the large model spread reported by Hill et al. [94], for which further inter-model comparison would be needed to more exactly attribute the discrepancies. Nevertheless, Cleo does indeed behave as expected for a SDM and is in agreement with other microphysics models for warm rain within the KiD framework.

B.7 CONCLUSIONS

Cleo is a concise and fully-functioning SDM for warm-cloud microphysics. We have validated our SDM against known test-cases for condensation/evaporation, collisions between droplets, and Lagrangian particle motion, including an integrated test including all three using the 1-D KiD framework. The motion is less expensive than higher-order methods yet still preserves the flow-field, is non-diffusive, and has increasing accuracy

with increasing grid resolution. For condensation/evaporation we solve the Köhler theory ODE explicitly. Our method incorporates ventilation effects and droplet (re)activation as well as a number of cost-saving measures including adaptive sub-time-stepping. For collisions between droplets, Cleo has various options which reflect the uncertainty in the current knowledge about collision probabilities and their outcomes. Cleo can easily switch between different formulations of the collision kernel and, motivated by a need to investigate the possible influence of collisional breakup on rain evaporation and downdraughts, Cleo includes a framework to model collisional breakup and rebound. We do this in a way that still respects the key features of the original collision-coalescence algorithm by conserving superdroplet number and decreasing the variance in the collisions with decreasing superdroplet multiplicity.

Cleo is useful to a broad group of researchers that require a model for warm-cloud microphysics. Simulations can be up-to 3-D and can include any combination of the primary microphysical processes behind warm-rain, each with their own independent time-step. Cleo can be stand-alone, perform “piggybacking” of LES [80], or be used as a fully fledged microphysics scheme. Like other SDMs, Cleo provides a precise representation of the droplet size distribution and an intuitive depiction of microphysical processes. We have also made sensitivity studies easy to conduct through Cleo’s high-degree of flexibility, whereby microphysics and data output can be easily customised in order to assess the impact of individual microphysical processes [22], potentially for probing microphysical uncertainties. Finally, since Cleo is also well-suited to high-performance computers, it has the potential to be used in more demanding simulations which probe how cloud microphysics and its uncertainties propagate to larger spatial and temporal scales.

B.8 CODE AVAILABILITY

Cleo is published on it’s GitHub page: github.com/yoctoyotta1024/Cleo, alongside it’s documentation: <https://yoctoyotta1024.github.io/Cleo>. Version v0.52.0 is described and tested in this paper.

B.9 CODE AND DATA AVAILABILITY

All the code and results included this paper can be accessed from the dataset: Bayley [26].

B.10 APPENDIX

B.10.1 Example Parametrisation for Fragments From Collision Breakup

Here we expound one of Cleo’s options for how to determine the properties of a superdroplet which represents the fragments from collisional breakup; namely how to define ξ_{frag} , R_{frag} and $M_{s,\text{frag}}$ in Section B.4.2 for Equations (B.25) to (B.32). As a simple approach for exploring the potential importance of collisional breakup in warm-cloud microphysics, here we assert mass conservation and that the total number of fragments, N_{frag} , depends on a col-

lision's kinetic energy, T_E . Schlottke et al. [230] reported that N_{frag} depends on T_E as:

$$N_{\text{frag}} = \left(\frac{3}{2} - (T_E/\mu J)^{0.135} \right)^{-1}. \quad (\text{B.37})$$

Since this equation diverges at $T_E/\mu J = 1.5^{1/0.135}$, we limit $N_{\text{frag}} \leq 25$ for $T_E/\mu J > 16.50$; and so that collisional breakup always results in more droplets than collided, even when $\xi_j = \xi_k = 1$, we impose the condition that $N_{\text{frag}} \geq 2.5$. Both of these amendments are concordant with the experimental data presented by Schlottke et al. [230].

Given N_{frag} is the number of fragments produced by a single (real) droplet breakup event, the superdroplet multiplicity must be as follows,

$$\xi_{\text{frag}} = \text{round}(\xi_k * N_{\text{frag}}), \quad (\text{B.38})$$

and therefore, to ensure mass conservation,

$$R_{\text{frag}} = \left[\frac{\xi_k}{\xi_{\text{frag}}} (R_j^3 + R_k^3) \right]^{1/3}, \quad (\text{B.39})$$

and

$$M_{s,\text{frag}} = \left[\frac{\xi_k}{\xi_{\text{frag}}} (M_{s,j} + M_{s,k}) \right]. \quad (\text{B.40})$$

B.11 COMPETING INTERESTS

The contact author has declared that none of the authors has any competing interests.

B.12 ACKNOWLEDGEMENTS

C. J.A. Bayley thanks Sylwester Arabas (AGH University of Krakow, Poland) and Emma Ware (University of California, Davis, USA) for their engaging discussions and technical support setting up the 1-D KiD model using PyM-PDATA. We gratefully acknowledge code contributions to Cleo from Sergey Kosukhin and Lukas Kluft (Max Planck Institute for Meteorology, Germany; MPI-M). A special thanks is given to Yvonne Schrader (MPI-M) for her excellent advice on the graphic designs in this paper, as well as Romain Fiévet (MPI-M) for conducting the MPI-M internal review.

A. K. Naumann, R. Vogel and F. Poydenot have received funding which supported this work from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2037 "Climate, Climatic Change, and Society" (project number 390683824). R. Vogel further acknowledges support from an ERC starting grant (ROTOR, grant no. 101116282). This project has received funding from Horizon Europe programme under Grant Agreement No 101137680 via project CERTAINTY (Cloud-aERosol inTeractions & their impActs IN The earth sYstem). The authors further express their appreciation for the work of the developers of the free and open-source software which underlies Cleo, especially from the developers of Git, GitHub, Python, the C++ standard libraries, and, above all,

Kokkos. We also thank the Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG) from the information and communication services Cleo's development has benefited from, and finally we thank the Deutsche Klimarechenzentrum (DKRZ) for the computer facilities from project 1183 we used to conduct this work.

B.13 LARGE FIGURES

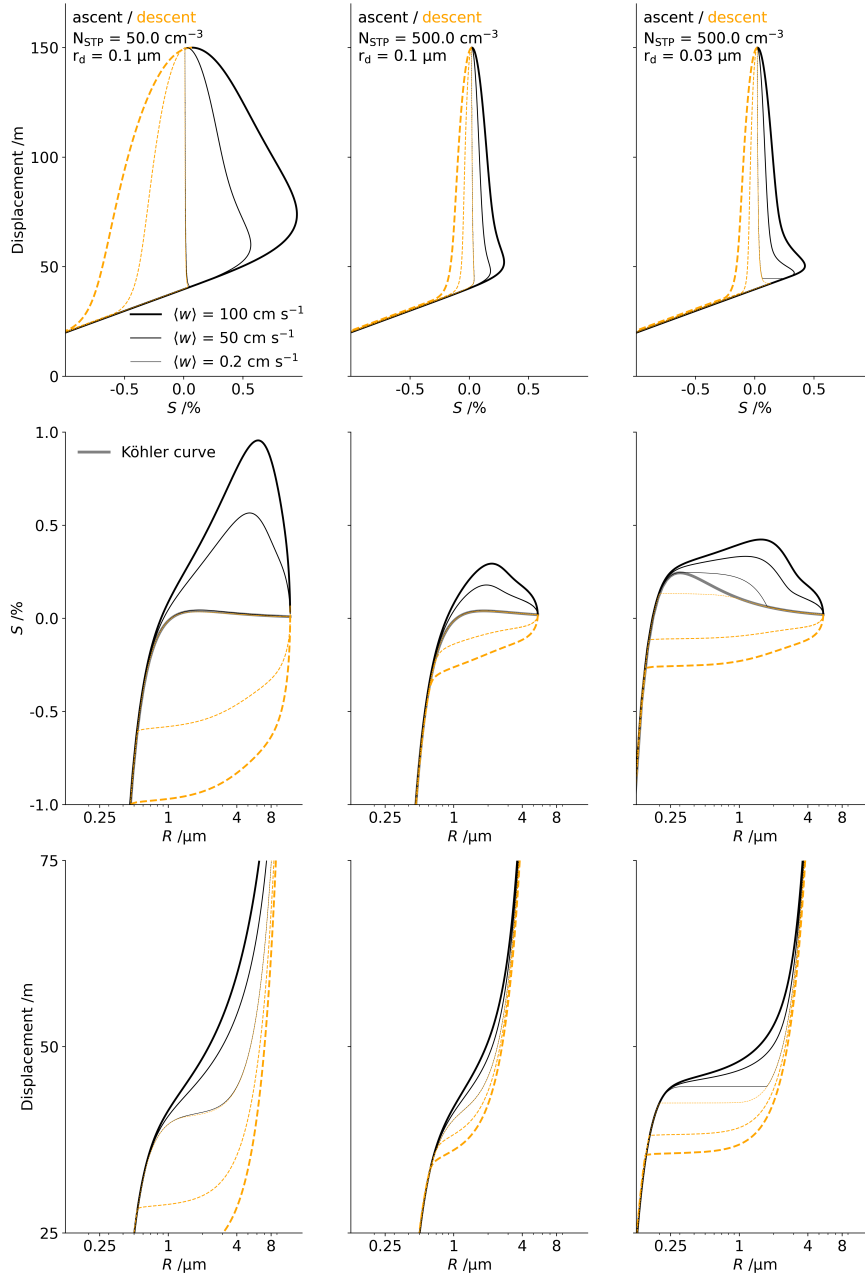


Figure B.9: Test of condensation/evaporation as in Arabas and Shima [13, Sect.7, Fig. 5]. The panels are the same as in Arabas and Shima [13], except the radius in the third column is $0.02 \mu\text{m}$ smaller, because slightly different numerics and initial thermodynamic conditions affect the radius at which cusp bifurcation is observed. Each column shows the evolution of an air parcel as it adiabatically rises and falls with a certain number concentration and initial radius of mono-distributed dry aerosol. Top row: the displacement of the parcel against its supersaturation; Middle row: The supersaturation as a function of the droplet radii; Bottom row: the displacement of the parcel against the droplet radii.

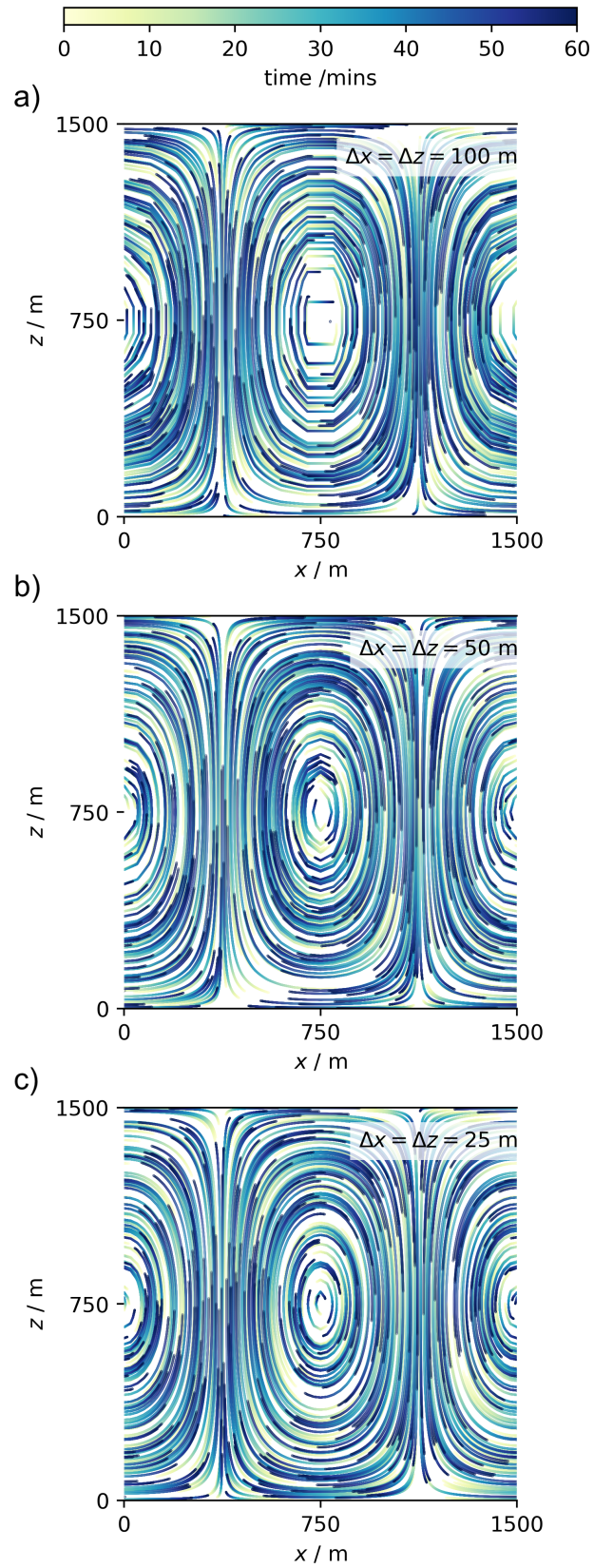


Figure B.10: Results of tracer particle motion in the 2-D divergence free laminar flow described by Arabas et al. [11]. Grid-spacing decreases from top to bottom panel.

CONSTRAINING RAIN EVAPORATION FROM SHALLOW-CLOUDS IN THE TRADES USING AN OBSERVATION-BASED SUPERDROPLET MODEL

The work in this appendix is currently under review at *Atmospheric Chemistry and Physics* as C. J. A. Bayley* and N. Niebaum* et al., ‘Constraining Rain Evaporation from Shallow-Clouds in the Trades using an Observation-Based Superdroplet Model’, *EGUsphere* **2025, 1–30 (2025).**

Authors: Clara J.A. Bayley^{*1,2}, Nils Niebaum^{*3}, Florian Poydenot³, Ann Kristin Naumann^{1,3,4}, Mampi Sarkar^{5,6}, and Raphaela Vogel³

^{*}These authors contributed equally to this work.

¹Max-Planck-Institut für Meteorologie, Hamburg, Germany

²International Max Planck Research School on Earth System Modelling, Hamburg, Germany

³Meteorologisches Institut, Universität Hamburg, Hamburg, Germany

⁴Ludwig-Maximilians-Universität München, Munich, Germany

⁵Institute of Climate and Atmospheric Science, University of Houston, Houston, TX, USA

⁶Department of Earth and Atmospheric Sciences, University of Houston, Houston, TX, USA

Author contributions: NN fitted the observational data, ran the model, and processed the output. CJAB was the main developer of Cleo for the project; she constructed the model setups and worked collaboratively with NN to interpret the results, and to write and edit the manuscript. RV and CJAB conceptualized the project, and RV oversaw the project’s development and had many discussions with CJAB and NN to discuss the observations, the results, and the manuscript. FP contributed the single-droplet analytical approximation to the results. AKN and MP provided scientific support, interpretation of results and feedback on the manuscript.

C.1 ABSTRACT

Rain evaporation influences subcloud moisture and energy budgets as well as boundary layer buoyancy, making it important for shallow cumulus organization. This study constrains the amount of, and controls on, rain evaporation in the subcloud layer of trade-wind shallow cumuli. We combine observations from the EUREC⁴A field campaign with a 1D rainshaft using superdroplet model (SDM) microphysics. With SDM we explicitly resolve how the observed droplet size distributions (DSDs) evolve beneath cloud-base, and explore, for the first time, the potential impact of raindrop collisions. Constraining the amount of evaporation, we find mean column integrated evaporative cooling of $150 \pm 170 \text{ W m}^{-2}$, and evaporation fraction of $30 \pm 20\%$. The dominant controls on rain evaporation are the subcloud relative humidity and the cloud-base DSD. Subcloud relative humidity determines the slope of the evaporation profiles, and cloud-base rain water content (RWC) controls the column integrated cooling. In contrast, evaporation fraction can vary substantially between clouds, even with comparable RWC, because of small differences to the cloud-base DSD. For narrow DSDs, the evaporation fraction is well captured by an analytical approximation, which also highlights the importance of ventilation effects on evaporation. We find collisions have negligible impact on rain evaporation, however, beyond rain rates of $\approx 3 \text{ mm h}^{-1}$, differences emerge which suggest that collisions, in particular those causing severe droplet breakup, need to be considered for heavier precipitation. Our results stress the importance of accurately measuring and modeling DSDs for rain evaporation, but justify the use of simpler microphysics models that only represent condensation/evaporation for lightly precipitating regimes as observed during EUREC⁴A.

C.2 INTRODUCTION

Below every precipitating cloud, raindrops pass through unsaturated air on their way towards the surface and evaporate either partially or completely. This sub-cloud layer rain evaporation reduces precipitation efficiency and therefore the amount of rain that reaches the surface [183, 184, 211, 222, 235]. Rain evaporation also moistens and cools the surrounding air, thereby influencing the energetics and stability of the atmosphere. But above all, rain evaporation is a key component of the cloud life cycle. The cooling effect from rain evaporation makes air negatively buoyant, driving downdrafts which spread radially outward when reaching the surface in the form of cold-pools [e.g. 216, 312]. These cold pools go on to trigger a chain of dynamic processes that influence the mesoscale organization and amount of clouds [7, 298, 313]. In the trade-wind regions, precipitating cumuli are frequent and, due to the relative dryness of the trades, rain evaporates efficiently and triggers cold pools frequently [193, 298]. Overall, cold pools occur about 7.8% of the time during a day and about 73% when averaged over all days in the winter trades [281, 298]. Cold pools thus have a strong impact on both the coverage and organization of trade cumuli, with consequences for cloud radiative feedbacks and precipitation patterns [7, 211, 212, 231, 298, 313]. Rain evaporation therefore has a critical impact on the energetics, structure, and cloudiness of the trade-wind boundary layer. However, its magnitude and controls have not yet been constrained.

Due to a shortage of field observations, it is difficult to robustly quantify rain evaporation rates from direct measurements. Tridon et al. [282] developed a method to obtain droplet size distributions (DSDs) from Doppler velocity spectra measured with K_a and W band Doppler radars. They retrieved evaporative cooling rates from a continental deep convection system in the sub-cloud layer between 0.75 – 3 km above sea level, with a large range up to 60 K d^{-1} (or equivalently roughly 422 W m^{-2}). Their retrieved vertical profiles indicate most evaporation in the middle of the sub-cloud layer. However there are large uncertainties in the retrieval method of Tridon et al. [282], and, because deep continental convection differs strongly from winter trades, direct analogies are limited. Disentangling the various factors influencing rain evaporation using field observations remains even more elusive.

Models for rain evaporation are also scarce, and so far limited to models of condensation/evaporation using bin microphysics schemes. Sarkar et al. [222] used DSDs from in-situ cloud probes during the EUREC⁴A/ATOMIC campaign [144, 201] to simulate trade-wind rain evaporation with a 1-D rainshaft model with bin scheme microphysics. They found 63% of the rain mass to evaporate in the sub-cloud layer in the mean, and a spread in latent cooling of $15 - 352 \text{ W m}^{-2}$. Unlike Tridon et al. [282], their evaporation profiles showed stronger evaporation near the cloud-base (top-heavy) than near the surface (bottom-heavy) for the majority of profiles. Sarkar et al. [222] were able to show the measured rain water content (RWC) was a dominant constraint on the column integrated evaporation rate. They also identified that the shape and mean radius of the DSD controls the top- vs. bottom-heaviness, with larger raindrops favouring bottom-heaviness. The large inter-cloud spread of evaporation simulated by Sarkar et al. [222] highlight the challenge in constraining the magnitude, variability and vertical structure of rain evaporation, for which we need both more observations and modeling.

An alternative and arguably more realistic way to model rain evaporation is through Superdroplet Models for cloud microphysics [SDMs; 9, 11, 21, 22, 36, 96, 187, 215, 241, 249]. SDM is a Lagrangian particle model which represents a group of similar droplets with comparable radius, velocity, and mass of solute through one simulated particle called a superdroplet. The multiplicity of a superdroplet accounts for the number of real droplets it represents, and all microphysical processes are modeled explicitly (i.e. each superdroplet evolves freely and individually), except for collisions between droplets which are modeled probabilistically [23, 241].

In contrast to classical bin and bulk schemes, SDM has numerous conceptual and computational advantages [78, 182]. Relevant for this study, are that SDM does not suffer from numerical diffusion, it makes no assumptions about droplet categories nor size distributions, and it models condensation/evaporation according to fundamental Köhler theory [134, 208]. The strong relationship between SDM and our fundamental theoretical understanding of cloud microphysics makes SDM a highly suitable model for studying the magnitude and controls on rain evaporation.

By using a SDM, we are also able to investigate the indirect effect of droplet collisions on rain evaporation. Previous studies using bin scheme microphysics neglected the effect of collisions between droplets on rain evaporation [222, 223]. However, whilst collisions between droplets do not change the total mass of raindrops, they do change the distribution of mass across

raindrop sizes, and therefore could impact the total rate of evaporation. Collisions leading to droplet coalescence shift mass towards larger droplets, which fall faster and have a smaller surface to volume ratio than smaller droplets, therefore lowering the evaporation rate [105, 236]. Conversely, collisional breakup shifts mass towards smaller, faster-evaporating droplets, and the higher the number of fragments, the stronger this effect. Rebound has no effect except that it makes coalescence less likely than if it did not occur. The net effect of collisions (coalescence, rebound, and breakup) on sub-cloud rain evaporation has not been assessed before, but nevertheless may be significant.

Taking into account the advantages of SDM and our ability to model droplet collisions with it, this study uses SDM to quantify rain evaporation beneath trade-wind shallow-cumuli, as well as elucidate how the evaporation is impacted by microphysical processes, ambient thermodynamic conditions (including cloud-base height), and the DSD. We setup the SDM called Cleo [22, 23] to model steady-state 1-D rainshafts for clouds with cloud-base DSDs and thermodynamic profiles fitted from the EUREC⁴A campaign. To our knowledge this is the first time SDM has been applied to quantify sub-cloud rain evaporation in the trades, also constrained by field observations, and accounting for the effect of droplet collisions.

The paper is structured as follows: Section C.3 describes how we used the observational data and the setup of our rainshaft model; Section C.4 presents the influence we have found of ambient thermodynamics, the cloud-base DSD, and microphysical processes on rain evaporation; Section C.5 summarizes our findings and gives an outlook.

C.3 DATA AND METHODS

C.3.1 *Observational Data*

To model realistic conditions representative of the sub-cloud layer beneath shallow trade-wind cumuli, we use observational datasets for the cloud-base DSDs and thermodynamic conditions from the EUREC⁴A campaign (Elucidating the Role of Clouds-Circulation Coupling in Climate) as the basis for our rainshaft. EUREC⁴A took place in the tropical North Atlantic upstream of Barbados from January 20th to February 20th 2020 to understand and quantify the interaction of trade cumulus clouds with their atmospheric and oceanographic environment across a range of scales, from micro- to meso-scale [35, 263].

For the cloud-base DSDs, we use in-situ measurements at 1 Hz resolution, corresponding to approximately 100 m spacing between measurements. We use the pre-merged product by Coutris [48] based on the CDP-2 and 2DS instruments onboard the ATR aircraft [32]. The measurements cover droplet radii from 1 μm to 1.275 mm in non-uniform radii bins [32, 48]. From this dataset, we identify individual “rain clouds” from 19 cloud-base rectangular flight tracks between 500 – 1200 m altitude flown east of Barbados over 11 days following the definition from Bony et al. [32] that rain measurements contain raindrops of at least 250 μm radius. We then define an individual cloud as at least three consecutive rain measurements, which gives all clouds a horizontal extent of at least 300 m, given the ATRs flight speed of approximately 100 m s^{-1} . Because the concentration of raindrops is very low compared to the measurement volume of the devices (30 cm^3), we allow for

up to five consecutive rain-free measurements in-between individual rain measurements to still count as one rain cloud.

To facilitate comparability to other studies we fit the cloud-base DSD measurements for each rain cloud with a bimodal log-normal distribution. Although this is not a necessary requirement for SDM, by fitting an analytic form to the cloud-base DSDs, bulk and bin schemes, which necessarily or often make similar approximations, can also use our droplet initialization data [e.g. 222, 236]. The bimodal log-normal DSD of each cloud is the sum of two log-normal probability distributions, $L(r)$, for droplets of radius r following Equation (C.1), where N_a is the total number of drops, μ_g is the geometric mean radius of the distribution (i.e. the median of the distribution), and σ_g is the geometric standard deviation of the distribution.

$$L(r) = \frac{N_a}{\sqrt{2\pi} r \ln \sigma_g} \exp\left(-\frac{(\ln r - \ln \mu_g)^2}{2(\ln \sigma_g)^2}\right) \quad \text{for } r > 0 \quad (\text{C.1})$$

We further choose to fit the bimodal distributions only for radii greater than $50 \mu\text{m}$, which represents the boundary to cloud droplets [32, 248]; thereby avoiding modeling droplets which do not fall at an appreciable speed and therefore are not relevant to the rainshaft. Our fitting algorithm also weights individual radii bins of the observational data based on the cube of the radius, giving greater weight to larger radii bins in the loss function. These would otherwise be underrepresented compared to the highly numerous small droplets, although they contain most of the RWC. As a final constraint on the modeled DSDs, we remove 13 outliers whose fitted distributions do not agree with the observed RWC (Section C.3.3).

For most clouds, the fitted bimodal log-normal distribution is well in line with the temporal mean of the observations. Figure C.14c shows the fitted DSD overlain on the measured droplet sizes for a typical cloud. The fitted cloud RWC in our model also aligns closely with the observed temporal mean RWC for each cloud, as shown by Figure C.14d, with a Pearson correlation coefficient $R = 0.99$ (log-log Pearson correlation coefficient of $R_{\log} = 0.91$).

We use the JOANNE dropsonde dataset Level 3 Version 2.0.0 [68, 69] to get humidity and temperature profiles in the vicinity of the measured DSDs for each cloud. First we identify the dropsondes released by HALO and P3 within 100 km and within ± 3 h of the DSD measurements made for each cloud. This definition is contentious, since the clouds whose DSDs we measure have much smaller time- and length-scales, but it allows for a more representative sample of the dropsonde measurements. After identifying corresponding dropsondes for individual clouds, we fit the thermodynamic profiles for each cloud with a constant potential temperature, a linear relative humidity, and a linear pressure profile. For the potential temperature and pressure profiles, we fit to the dropsonde measurements between 200 – 500 m. For the relative humidity profile, we also fit to the dropsonde measurements between 200 – 500 m but give dropsondes with moister profiles between 100 – 700 m more weight (using a smoothed weight with constant prior). We do so, because moister profiles are more likely to represent atmospheric conditions below clouds than the clear-sky regions. Finally, cloud-base is defined as the altitude of 100% relative humidity. An example for the thermodynamic fits can be seen in Figure C.14a,b,c. Figure C.14b further indicates that our fit follows the moister profiles more closely, and

that the cloud-base altitude at 100% relative humidity aligns well with the ATR flight altitude.

The resulting dataset excluding outliers contains 102 precipitating trade-wind shallow-cumuli clouds, each with their own fitted thermodynamic profiles and DSDs.

c.3.2 *Setup of the Superdroplet Model Cleo*

We use the SDM Cleo v0.39.7 to simulate all the relevant warm-cloud microphysical processes for our study [22, 28]. We run Cleo as a quasi-stationary 1-D rainshaft to represent each cloud in our sample of identified rain clouds from the EUREC⁴A observations (Section C.3.1). This means we keep the thermodynamic profiles and the cloud-base DSD constant in time, matched to the observations, and examine the time-averaged statistics of the sub-cloud layer once it reaches a quasi-stationary state (Section C.3.3). In this way we obtain a “snapshot” of each cloud’s precipitation flux, evaporation rate and evaporation fraction, but do not explore the time-evolution of the cloud, in which heavy rain may influence its environment e.g. through turbulent mixing, downdrafts, and aerosol depletion.

For each cloud in our sample, we divide the 1-D rainshaft into a cloud-base and a sub-cloud layer. The cloud-base is a single layer 100 m thick with constant thermodynamics, and where the fitted DSD is applied by randomly sampling it at every motion time-step (2 s). To randomly sample the DSD, we randomly draw a superdroplet from each of 2048 uniformly-spaced logarithmic bins between 50 μm to 3 mm and assign its multiplicity according to the fitted DSD. At the end of each timestep we remove any remaining superdroplets from the cloud-base layer and initialize 2048 new superdroplets. Superdroplets which fall into the (initially empty) sub-cloud layer are not removed and thereafter evolve freely. The sub-cloud layer has a 20 m grid-spacing and the total height of each column varies with the fitted cloud-base height. Within the quasi-stationary state each gridbox contains roughly 128 activated superdroplets, which is deemed sufficient for convergence of the SDM collision algorithm, given we average over the quasi-stationary state (equivalent to approximately 1000 separate SDM realizations) [177, 182, 286, 287].

We use four microphysical setups to investigate the importance of droplet collisions, and these are summarized in Table C.1. All the setups simulate droplet motion (droplets falling according to their terminal velocity) and evaporation/condensation. The EvapOnly setup models no additional processes and is considered as the “control” scenario because it is most comparable to previous studies which also quantified rain evaporation [222, 223, 282]. The EvapCoal setup additionally simulates collision-coalescence to examine the effect of collisions between droplets in the absence of collisional breakup and rebound, as is typically assumed by conventional microphysics models. The EvapCoalBuRe-many and EvapCoalBuRe-few setups further include collisional rebound and breakup as well as coalescence and they differ in the fragments they produce when breakup occurs. Due to the large uncertainty in parameterizing fragments from collisional breakup, we choose to prescribe a fixed number of fragments per breakup event: EvapCoalBuRe-few produces 5, whereas EvapCoalBuRe-many produces 50. EvapCoalBuRe-few (5 fragments) represents the lower bound and EvapCoalBuRe-many (50 fragments) represents the upper bound on evaporation enhancement

by breakup. This bracketing strategy bounds sensitivity of evaporation to breakup while ensuring that our results are not tied to a single uncertain parameterisation.

Cleo outputs evaporation rates directly from the Köhler theory ODE for the change in radii of the droplets via evaporation/condensation [23]. The column integrated evaporation mass flux, $F_{e,tot}$, is then the sum of the evaporation rate across all gridboxes, and the evaporation fraction (ξ) is given as the fraction of quasi-stationary mean integrated evaporation flux and the quasi-stationary mean cloud-base precipitation flux ($F_{p,cb}$), thus $\xi = \langle F_{e,tot} \rangle / \langle F_{p,cb} \rangle$ (Section C.3.3).

In the following sections, the mean mass radius, $\bar{r}_M$, is the mean of the mass weighted distribution of droplet radii. It is calculated as $\bar{r}_M = \sum_i \frac{r_i \cdot m_i}{M}$, where r_i and m_i are the radius and mass of an individual raindrop ($i \in [0, N]$) and M is the total mass of all raindrops, $M = \sum_i m_i$.

C.3.3 Validation of model setup

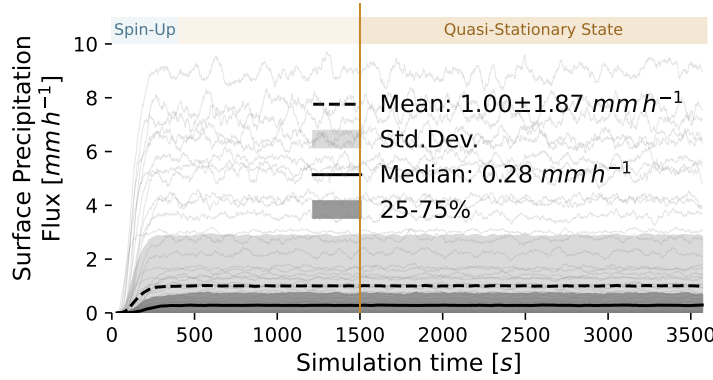


Figure C.1: Temporal evolution of surface precipitation in the EvapOnly setup against simulation time. Individual clouds are shown as thin, grey lines with a rolling mean over 60s. The mean over all clouds is shown as a dashed thick black line with the corresponding standard deviation in light grey shading. The median over the clouds as solid thick black line, with the inter-quartile-range as dark grey shading. Colored horizontal bars at the top roughly indicate the spin-up phase in blue and quasi-stationary state in orange. We use data past 1500s (vertical orange line) as quasi-stationary state for our analysis. Temporal mean values over the quasi-stationary state are shown in the legend.

Quasi-stationary state: The spin-up phase, when droplets initially start falling into the sub-cloud layer and precipitation increases, lasts about 300s (Figure C.1). After the spin-up, all clouds show stable precipitation and the model is in a quasi-stationary state with variance induced by our random sampling of superdroplets. We analyse the mean over the quasi-stationary state from 1500s until the end of the simulation at 3600s, thereby ensuring that even the slowest falling droplets have penetrated into the surface layers, and so the DSD at each height is in dynamic equilibrium. In this way we

obtain our “snapshots” of each cloud’s precipitation flux, evaporation rate and evaporation fraction.

Outliers: We omit all clouds which exceed the inter-cloud mean surface precipitation and column integrated evaporation rate by more than 4 standard deviations. In other words, we omit 1.7% of the clouds, those that have cloud-base precipitation exceeding 23.18 mm h^{-1} and column integrated evaporation of 2.37 mm h^{-1} . We further exclude 13 clouds for which the fitted DSD’s RWC is not within 66 – 150% of the observed RWC. In total, we analyze 102 clouds.

Precipitation and RWC: Figure C.2 shows the distributions of cloud RWC and surface precipitation flux for the quasi-stationary state of all the clouds in the EvapOnly simulations. The distribution of surface precipitation flux covers a range from 0 – 10 mm h^{-1} with a strong skew towards lower values, a mean of $1.00 \pm 1.89 \text{ mm h}^{-1}$ and a median of 0.28 mm h^{-1} . This spread is also aligned with observations, such as those taken aboard of *MS Merian* during the EUREC⁴A campaign, which measured mean surface rain rates ranging from 0 – 7.43 mm h^{-1} [4]. In addition, our values span a range which includes both the climatological mean rain rates and rain intensity of trade wind cold pools identified by Vogel et al. [298] at the Barbados Cloud Observatory.

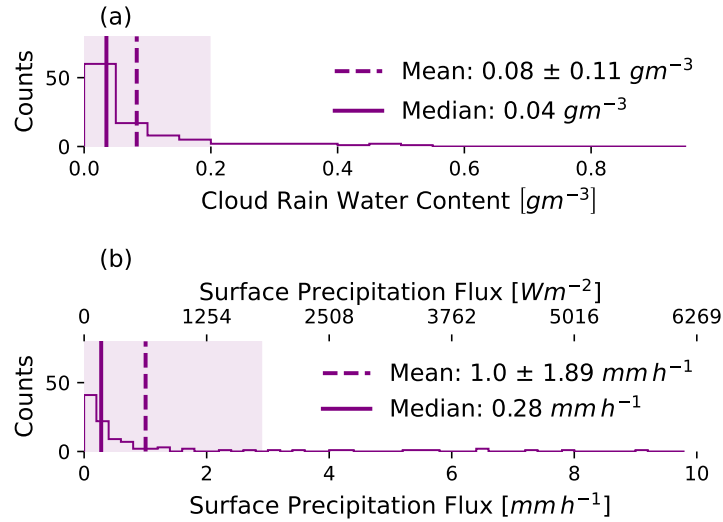


Figure C.2: Histograms of (a) cloud RWC and (b) surface precipitation flux for the EvapOnly setup. Mean denoted as dashed vertical line with standard deviation as shading. Median as solid vertical line. Values are shown in the legend.

C.4 RESULTS AND DISCUSSION

First of all, we quantify the sub-cloud evaporation for our sample of 102 clouds simulated with the “control”, EvapOnly setup. Figure C.3 shows the distributions of column integrated evaporative cooling and evaporation fraction and both distributions are skewed towards lower values: the evaporative cooling ranges from 0 – 800 W m^{-2} with a mean of $150 \pm 170 \text{ W m}^{-2}$ and a median of 72 W m^{-2} . The evaporation fraction ranges from 0 – 80% and

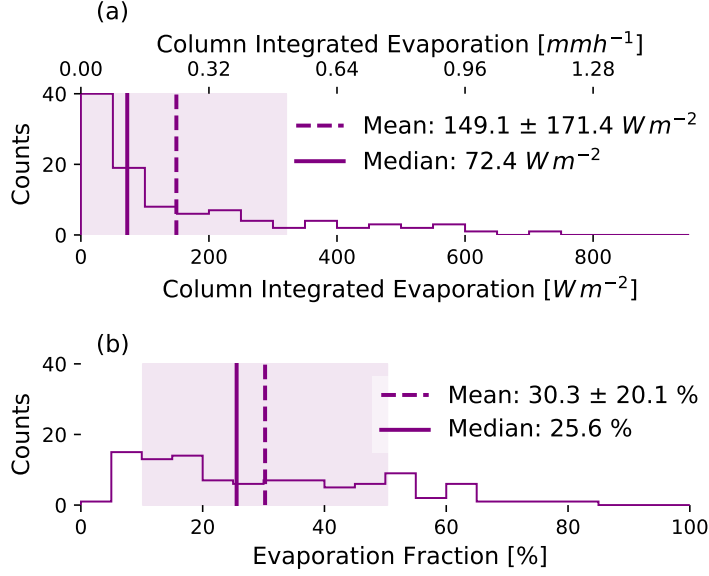


Figure C.3: Histograms of (a) column integrated evaporative cooling and (b) evaporation fraction for the EvapOnly setup. Mean denoted as dashed vertical line with standard deviation as shading. Median as solid vertical line. Values are shown in the legend.

is less skewed, with a mean of $30 \pm 20\%$ and a median of 25.6% . Somewhat surprisingly, the column integrated evaporative cooling and the evaporation fraction are independent quantities, with a correlation $R = 0.00$ (Figure C.11a).

The vertical evaporation profile is remarkably similar for almost all the clouds, with more evaporation near the surface (bottom-heavy). Figure C.4 shows the vertical profile of evaporative cooling with altitude, normalized to the individual cloud-base height for each cloud. Whilst there is considerable spread in the evaporation rates, except for a few rare cases, the maximum evaporation rate occurs in the lowest third of the rainshaft for all the clouds.

We have varying degrees of agreement with the shallow cumuli modeled by Sarkar et al. [222]. Our column integrated evaporative cooling rates are higher than, but comparable to, the results of Sarkar et al. [222], who identified a range of $15 - 352 W m^{-2}$ (comparable also with various measurement campaigns they mention, around $100 W m^{-2}$). However, our estimated evaporation fraction is much lower than their average of 63% . Accordingly, whilst we observe bottom-heavy evaporation profiles far more frequently than top-heavy ones, Sarkar et al. [222] reported that top-heavy profiles were more common. We discuss plausible reasons for these discrepancies in Section C.4.1.

C.4.1 What controls the evaporation profile, rate and fraction?

In this section, we explore potential factors shaping sub-cloud rain evaporation focusing on the role of cloud-base precipitation, the cloud-base droplet size distribution and the sub-cloud layer relative humidity profile. Cloud-base precipitation is strongly correlated to cloud-base RWC via the droplets'

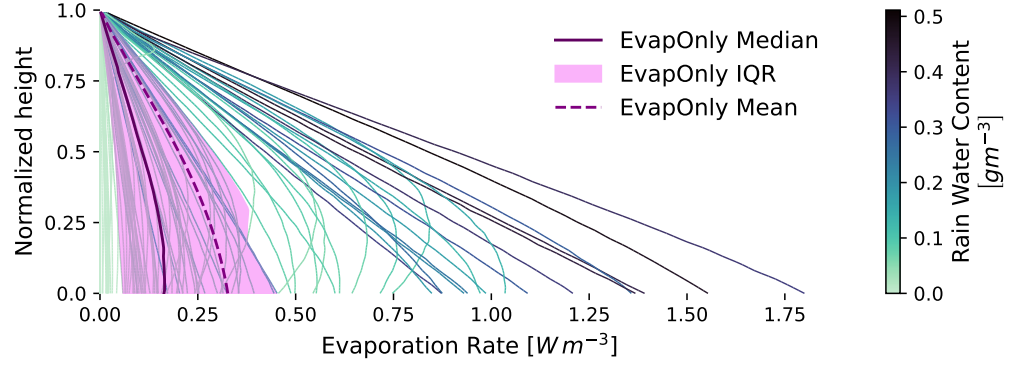


Figure C.4: Profile of the evaporative cooling rate against height for each cloud in the EvapOnly setup, normalized by each cloud's cloud-base height and colored by rain water content at the corresponding normalized height (thin lines). The mean (dashed purple line), median (solid dark purple line) and IQR (purple shading) are also shown.

fall speed, and so the results presented here for cloud-base RWC qualitatively also hold for cloud-base precipitation. The cloud-base DSD is well characterized by its mean mass radius, $\bar{r}_M$. Therefore, we begin by investigating how the evaporative cooling rate and the evaporation fraction of our clouds are related to the relative humidity, cloud-base RWC and $\bar{r}_M$. We present the results for the EvapOnly, "control", setup, and examine the effect of collisions between droplets afterwards in Section C.4.2.

To first order cloud-base RWC controls the amount of sub-cloud rain evaporation. Figure C.4 shows that below all clouds, RWC decreases towards the surface due to evaporation. Moreover clouds with higher cloud-base RWC, i.e. higher cloud-base precipitation flux, induce more evaporation, as seen both in the vertical profiles in Figure C.4 and integrated over the sub-cloud layer in Figure C.5 (see also Figure C.6a; correlation of $R = 0.90$ between cloud-base RWC and column integrated evaporative cooling). The higher the RWC at cloud-base, the more rain mass is available to evaporate in the sub-cloud layer. The remaining variance in evaporation can be explained by the diversity of DSDs and relative humidity profiles in our cloud sample.

The vertical gradient of evaporation is mainly ruled by the relative humidity profile, outweighing the effects of RWC decreasing towards the surface and the slight increase in $\bar{r}_M$. Figure C.7 shows a strong negative correlation of the evaporative cooling rate against relative humidity with height for nearly all clouds, with an inter-cloud mean Pearson correlation coefficient of -0.94 ± 0.13 . This is in line with expectations, that lower relative humidity leads to faster evaporation rates. Whilst RWC reduces towards the surface and so could cause a concurrent reduction in the evaporation rate towards the surface, we find that their correlation is negative -0.88 ± 0.16 , indicating that RWC responds to rather than controls the rate of evaporation. $\bar{r}_M$ typically increases towards the surface as the smaller droplets evaporate completely but the correlation between the vertical profile of evaporation rate and $\bar{r}_M$ is only slightly positive and highly variable ($R = 0.15 \pm 0.77$), because relative humidity is the more dominant control on the vertical gradient of rain evaporation.

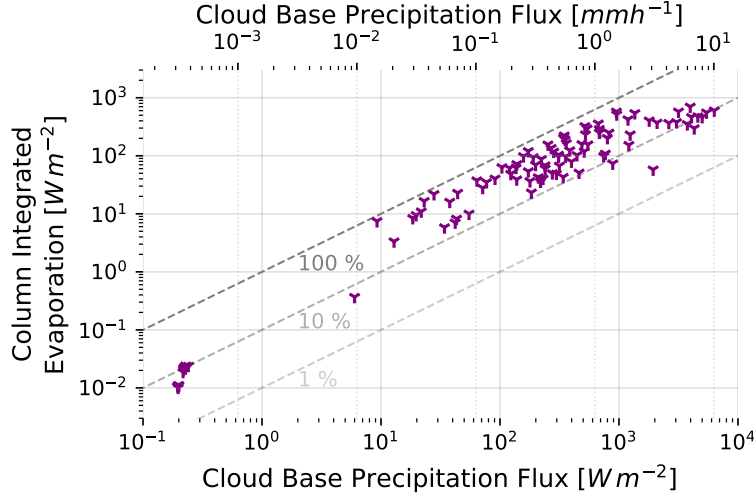


Figure C.5: Column integrated evaporative cooling against cloud-base precipitation flux for the EvapOnly setup. Dashed lines indicate contours of evaporation fraction of 100, 10, 1%.

Whilst the vertical profile of relative humidity shapes the profile of evaporation, the relative humidity does not appear to impact the column integrated (i.e. total) evaporative cooling strongly ($R = 0.01$; Figure C.13i). This may be because the fitted thermodynamic profiles do not vary sufficiently enough between clouds to cause appreciable differences in the total evaporation rate. We expect that for profiles in much more diverse environments, the impact of mean relative humidity would be greater, e.g. as shown by Srivastava [250, 251].

The mean evaporation height, a measure of the top-/bottom-heaviness of the evaporation profile, depends most strongly on $\bar{r}_M$ (Figure C.13k). The correlation between $\bar{r}_M$ and the mean evaporation height is negative for small $\bar{r}_M$ because increasing $\bar{r}_M$ in this regime reduces the number of droplets which completely evaporate higher in the column. Therefore clouds with smaller $\bar{r}_M$ have a more curved, top-heavy evaporation profile, similar to the profiles found by Sarkar et al. [222]. Meanwhile for larger $\bar{r}_M$, above about $750 \mu\text{m}$, the mean evaporation height shows little correlation with $\bar{r}_M$ since so few droplets completely evaporate. We also find very little impact of the cloud-base height on the evaporation fraction and column integrated evaporation (Figure C.13e,j,o).

The evaporation fraction is fundamentally ruled by raindrop size distribution at cloud-base. Higher $\bar{r}_M$ implies larger droplets in the cloud-base DSD, which evaporate less efficiently and thus lower the evaporation fraction. Thus, for the same RWC, a cloud with smaller $\bar{r}_M$ will have a higher evaporation fraction. Figure C.6b shows this relation of the evaporation fraction to $\bar{r}_M$, albeit the Pearson correlation coefficient underestimates the non-linear relation of the two variables ($R = -0.73$). Figure C.8 shows the same relation as well as an analytical single-droplet, mean-field approximation for the evaporation fraction. This approximation predicts that the evaporation fraction of a single droplet falling through a rainshaft with mean relative humidity and temperature scales with its radius as $r^{-7/4}$ (Appendix

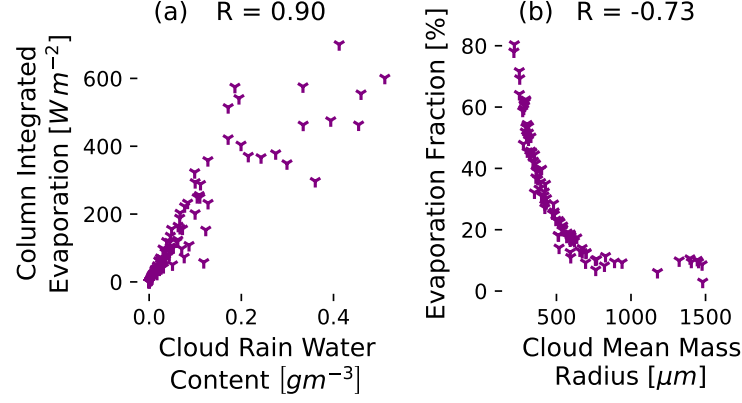


Figure C.6: The leading controls on sub-cloud layer evaporation for our sample of clouds, and their Pearson correlation coefficients. Each marker indicates an individual cloud. (a) Column integrated evaporative cooling is predominately controlled by the cloud-base RWC; (b) Evaporation fraction is predominately controlled by the cloud-base DSD, well-captured by the cloud-base mean mass radius, $\bar{r}_M$. (Additional correlations are also shown in Figure C.13.)

C.8.1). The approximation is well in line with the results of our SDM for most clouds with $\bar{r}_M$ between 150 – 1000 μm. The exceptions are clouds with broad DSDs, which, as expected, cannot be approximated well by a single radius and whose shape changes significantly with height. The spread amongst clouds with similar $\bar{r}_M$ but which still comply with the approximation can be explained by their slightly different relative humidity profiles and DSD widths. Since these differences are very small, the mean mass radius $\bar{r}_M$ of a DSD can be used to estimate the evaporation fraction well.

Figure C.8 also shows that accounting for the ventilation effect (the increase in evaporation rate due to droplet motion; Bayley et al. [23] and Pruppacher and Klett [208]) is crucial for modeling evaporation. Ventilation increases the evaporation rate of large raindrops significantly — roughly by a factor 9 for radii of 1 mm, and so omitting ventilation strongly reduces evaporation for all the clouds, with the strongest reduction for clouds with high $\bar{r}_M$. The size dependence of the evaporation fraction is proportional to the ventilation coefficient f_v (Appendix C.8.1). As such, setting $f_v = 1$ changes the evaporation fraction to scale with $r^{-5/2}$ instead of $r^{-7/4}$. Neglecting the ventilation effect, as is often done in microphysics models, therefore has serious implications for sub-cloud energetics and the formation of simulated cold pools.

The lower evaporation fraction, higher column integrated evaporative cooling, and more bottom-heavy vertical profiles that we have found compared to Sarkar et al. [222], are likely explained by our observations being downstream of theirs. Further downstream in the trades rain tends to be more abundant and stronger, resulting in measurements with higher RWC and larger $\bar{r}_M$. Sarkar et al. [222] prescribed cloud-base DSDs with a $\bar{r}_M$ as low as 0.1 mm, whilst $\bar{r}_M$ of our distributions lies between 0.2 – 2 mm. Therefore, we expect our simulations to show lower evaporation fractions and more bottom-heavy evaporation. In addition, the higher RWC of our clouds

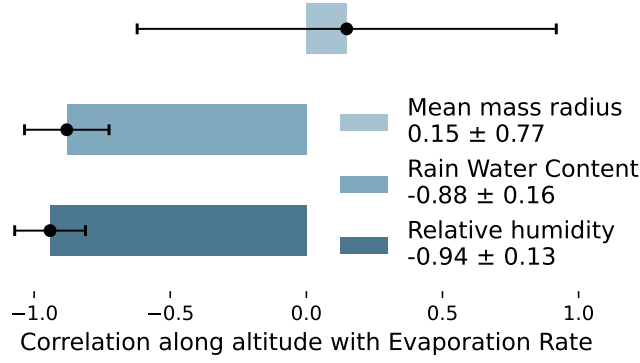


Figure C.7: Histogram shows the intercloud mean Pearson correlation coefficients along altitude between evaporation rate and relative humidity (dark blue), RWC (blue) and $\bar{r}_M$ (light blue) based on all simulated clouds for the EvapOnly setup. Respective inter-cloud mean and standard deviation of Pearson correlation coefficients shown in the legend.

leads to higher column integrated evaporative cooling. These differences highlight the major significance of cloud DSDs and their approximations for estimating evaporation rates in the sub-cloud layer of shallow-clouds.

c.4.2 Effect of Collisions Between Droplets

Collisions between raindrops induce systematic but only minor differences in rain evaporation for the clouds studied here. Figure C.9 shows the relative differences of the evaporation rate profiles between the setups accounting for collisions and the EvapOnly setup. The percentile shading also indicates a modest spread across the clouds simulated, which increases towards the surface because the differences induced at higher altitudes are propagated downwards and thus accumulate through the rainshaft.

Including collision-coalescence (EvapCoal) decreases the mean and median by up to 5% near the surface compared to the evaporation/condensation-only setup (Figure C.9a). However, including breakup and rebound as well as coalescence in the EvapCoalBuRe-few setup, opposes the coalescence effect and leads to a mean profile similar to the EvapOnly setup (Figure C.9b). If anything, the slight negative skew of the change in evaporation profiles compared to EvapOnly shows the combined effect of breakup, rebound and coalescence is more likely to increase evaporation for a population of clouds, however, for individual clouds it is not clear-cut whether the combined effect leads to more or less evaporation. It depends on the DSD, because the DSD controls the probability with which these three outcomes occur. Collisional breakup is more likely in broader distributions with larger $\bar{r}_M$, and this favors the increase of evaporation for these clouds because it creates more, and smaller droplets. Vice versa for narrower DSDs with smaller $\bar{r}_M$ whereby collision-coalescence is favoured which produces larger droplets that evaporate less efficiently.

Except for in cases of extreme breakup, the variability among our sample of shallow cumuli exceeds the impact of collisions on evaporation. The mean

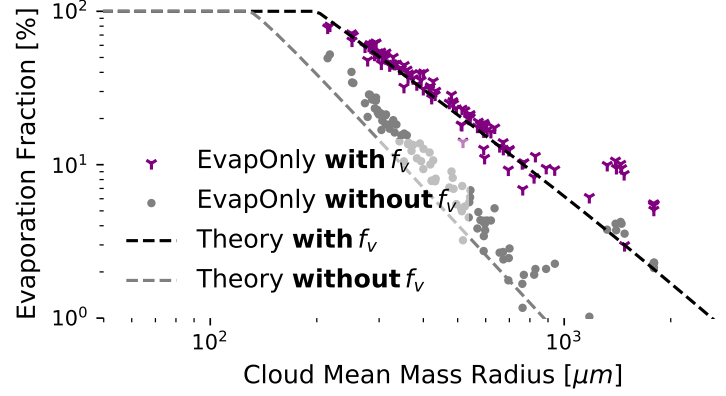


Figure C.8: Evaporation fraction against cloud-base $\overline{r_M}$ for all clouds in the EvapOnly setup (purple tri-blade markers). The Dashed black line shows the single-droplet analytical approximation for Köhler theory accounting for ventilation effects. The gray dots and gray dashed line show the same simulations and theory but neglecting the ventilation effect.

and standard deviation of the evaporation fraction and column integrated evaporative cooling are nearly identical for EvapCoal and EvapCoalBuRe-few setups, compared to EvapOnly values of $30 \pm 20\%$ and $150 \pm 170 \text{ W m}^{-2}$ respectively (Figure C.11a). The values for the EvapCoalBuRe-many setup, at $220 \pm 360 \text{ W m}^{-2}$ and $32 \pm 18\%$, are higher, because a few clouds undergo very extreme breakup in the sub-cloud layer which enhances evaporation there (Figure C.11a).

The lower tail of the distribution of vertical evaporation rates does not change when more fragments are produced during breakup (EvapCoalBuRe-many compared to EvapCoalBuRe-few), but the upper tail increases substantially (Figure C.9c). The upper tail contains the few clouds for which breakup occurs frequently and here we see increases the mean evaporation rate in the sub-cloud layer by 70% relative to the evaporation/condensation-only setup. Since EvapCoalBuRe-many drastically increases the number of fragments, to the extreme, these results can be considered an upper bound to the effect of breakup on evaporation rates beneath shallow cumuli in a more realistic setting.

Larger cloud-base precipitation fluxes lead to larger differences between the setups which account for collisions and the setup which only accounts for evaporation/condensation, differences both in evaporation fraction and column integrated evaporative cooling. In general, clouds with cloud-base precipitation above 2000 W m^{-2} , roughly corresponding to 3 mm h^{-1} and 1 g m^{-3} RWC, are more affected by collisions both in their column integrated evaporative cooling and in their evaporation fraction (Figure C.10, also Figure C.12a, c), especially in cases of extreme breakup. These results suggest that, although droplet collisions negligibly impact evaporation beneath the shallow trade-wind cumuli we have simulated, it's role becomes increasingly important as clouds deepen and their DSDs widen. We thus expect droplet collisions to play a greater role in controlling evaporation rates for more heavily precipitating clouds.

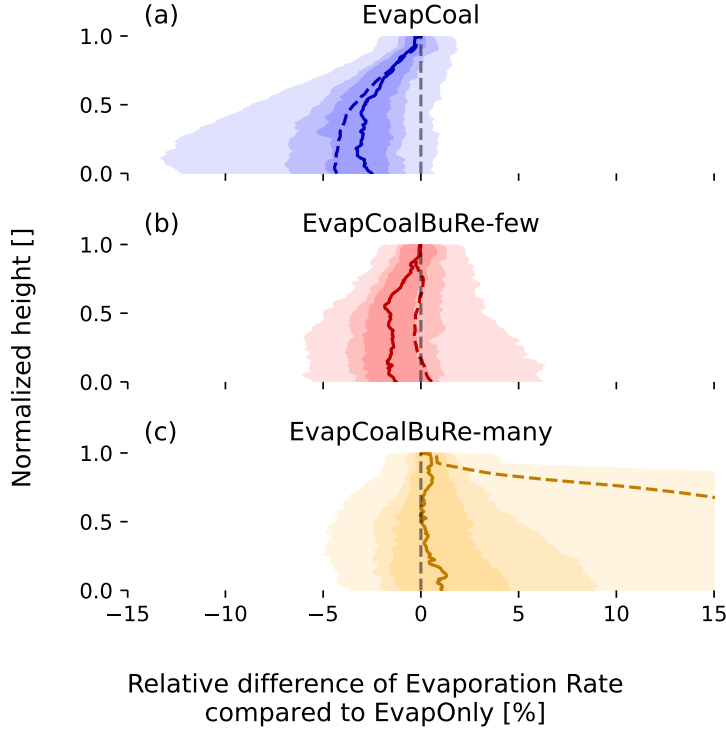


Figure C.9: Relative difference of evaporation profiles of (a) EvapCoal, (b) EvapCoalBuRe-few, (c) EvapCoalBuRe-many, compared to EvapOnly setup. Inter-cloud mean as dashed line. Inter-cloud median as solid line. Percentile range as shadings from dark to light are 33 – 66%, 25 – 75% and 10 – 90%.

C.5 SUMMARY AND OUTLOOK

C.5.1 Summary

Bridging the gap between field data and modeling, we use a superdroplet model (SDM) for cloud microphysics in conjunction with observations from the EUREC⁴A field campaign to estimate the strength of, and controls on, evaporation in the sub-cloud layer of precipitating trade-wind shallow cumuli. We identified 102 distinct clouds from the observations, with cloud-base DSDs from in-situ measurements and thermodynamic profiles from dropsonde measurements and used them to setup a steady-state 1-D rainshaft model so that we could analyse the evaporative cooling and evaporation fraction in the sub-cloud layer, in a way that is more directly comparable to observations than idealized studies. By using SDM, we represent all droplets without the need for arbitrary condensate categorization or any assumption on the droplet size distribution in the sub-cloud layer. Furthermore, we have a clear link between our model and the fundamental theories of cloud microphysics, meaning we can apply Köhler theory for condensation/evaporation explicitly, without being confounded by numerical diffusion, and we can test the impact of collision coalescence, rebound and breakup on rain evaporation for the first time.

Simulating only evaporation/condensation and droplet motion we estimate the column integrated evaporative cooling for our sample of clouds

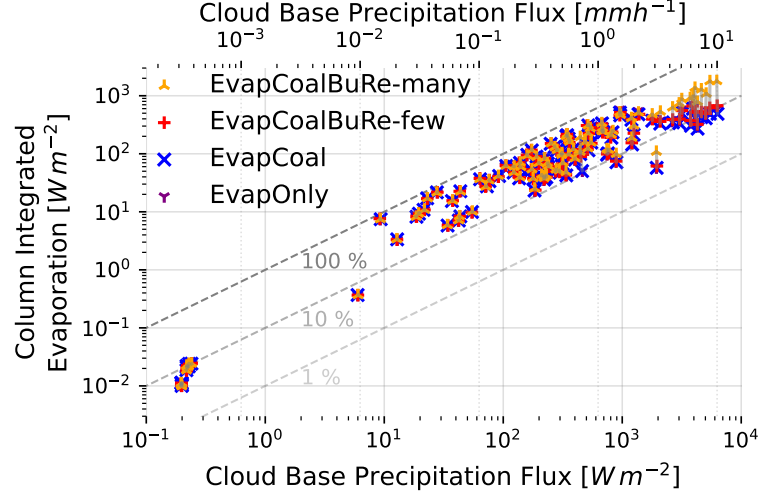


Figure C.10: As Figure C.5 but showing results for all microphysical setups with markers and colors as indicated in the legend. Dashed lines indicate contours of evaporation fraction of 100, 10, 1%. The four setups for an individual cloud are connected by a thin grey line.

ranges from 0 – 800 W m^{-2} , with a mean of $150 \pm 170 \text{ W m}^{-2}$ and a median of 72 W m^{-2} . Evaporated fraction ranges from 0 – 80% and is less skewed, with a mean of $30 \pm 20\%$ and a median of 25.6%. These values are in the same range as those of [222, 223], who also estimated rain evaporation from precipitating shallow-clouds, but with a bin microphysics scheme based on observations from another aircraft of the EUREC⁴A/ATOMIC campaign.

Based on our cloud sample we identified three key controls on rain evaporation in the sub-cloud layer. Cloud-base rain water content (RWC) rules the column integrated evaporative cooling and explains most of the variance across our sample ($R = 0.90$), in agreement with Sarkar et al. [222] and Sarkar et al. [223]. Relative humidity determines the overall gradient of the evaporation profiles with a mean correlation along altitude of $R = -0.94 \pm 0.13$. Our results show more bottom-heavy evaporation than Sarkar et al. [222], which can be explained by our DSDs having larger mean mass radii, $\bar{r}_M$. The role of $\bar{r}_M$ in shaping the top-/bottom-heaviness of the profiles also aligns with our finding that $\bar{r}_M$ is the key control on the evaporation fraction — as expected because larger droplets evaporate less efficiently. For narrow DSDs we can also match our evaporation fraction with a single-droplet mean-field approximation that has an analytical form.

For the first time, the effect of droplet collisions: coalescence, breakup, and rebound, on sub-cloud rain evaporation could be estimated. By using SDM we run all 102 cloud simulations under four different microphysical scenarios: excluding collisions (EvapOnly), including additionally collision-coalescence (EvapCoal), and further including rebound and breakup with either very few, or very many fragments produced in the event of breakup (EvapCoalBuRe-few and EvapCoalBuRe-many respectively). Including collisions in any scenario played a negligible role in our precipitating shallow-clouds sample, meaning that under light rain conditions, simplified micro-

physics which only considers condensation/evaporation is justified, as long as it accounts for ventilation effects. Nevertheless, we cannot rule out the importance of droplet collisions for more heavily precipitating clouds that have more than 2000 W m^{-2} cloud-base precipitation, roughly corresponding to 3 mm h^{-1} and 1 gm^{-3} RWC. For these clouds, the reduction in evaporation may easily exceed 10% if coalescence is included. Including breakup and rebound additionally reduces the impact of coalescence, and may even enhance rain evaporation, depending on the details of collisional breakup.

c.5.2 Outlook

Our results point to the importance of the DSD on evaporation beneath shallow-clouds. Clouds with similar cloud-base precipitation can produce different evaporation profiles solely due to differences in the shape of the DSD. The most likely reason for our lower evaporation fraction, higher column integrated evaporative cooling, and more bottom-heavy vertical profiles which we observe compared to Sarkar et al. [222] is the higher RWC and larger $\bar{r}_M$ of our cloud-base distributions. This points to the importance of accurately representing DSDs and their evolution, which requires not just a highly resolved microphysics scheme (SDMs, bin schemes, or many-moment bulk schemes), but also accurate observations and a suitable fit to model them. In the future, assuming the functional form of the DSD we use in models should also be avoided, since how they distort the observed DSD likely distorts the evaporation that is modeled too [109, 110].

It is encouraging that collisions play only a minor role in controlling evaporation for our cloud sample, but it is still unclear how heavily precipitating cases are affected. Since microphysics schemes tend to agree on the evolution of the DSD in the absence of collisions [94], it should be possible to achieve a broad consensus for evaporation rates beneath shallow-clouds, even including cruder models, as we show with our single-droplet analytical approximation (accounting for ventilation effects). What remains to be seen is whether this is true for a wider set of environmental and microphysical conditions than our sample allowed, especially more heavily precipitating cases. For this we require more long-term observations, which sample a more variable cloud population. We also require a clearer understanding of how to model droplet breakup because the uncertainties in the fragment size distribution prevent us from tightly constraining breakup's effect on rain evaporation. For heavier precipitation than we have sampled, models may have to account not only for the feedback of microphysics onto the relative humidity profile, which could amplify the role of collisions on evaporation rates, but also the impact of evaporation on turbulent mixing and buoyancy; both of which a non-steady-state rainshaft could capture.

Whilst we only studied precipitating trade-wind shallow cumuli observed during EUREC⁴A, our model is ready to be applied to any analogous observational datasets. This opens the door to its use on long-term measurements and to comparison with radar-based retrievals of rain evaporation [282]. A larger record of rain evaporation, which involves more clouds also under different environmental conditions, is needed to confirm that differences in our cloud samples cause the differences between our results and those of Sarkar et al. [222]. Such a record could also be used to extend our analysis to different cloud regimes. Measurements from the Barbados Cloud Observatory [258] are well-suited to this task, as they span the range of shallow and

deep convection typical of the dry and wet seasons in the tropics. Understanding and accurately modeling rain evaporation is crucial for predicting cold pool formation and cloud cover in the trades; our study provides new constraints on this process and in conjunction with further observations can improve how models represent shallow convection and its feedbacks in the climate system.

C.6 CODE AVAILABILITY

Cleo is published on it's GitHub page: github.com/yoctoyotta1024/Cleo, alongside it's documentation: <https://yoctoyotta1024.github.io/Cleo>. Version v0.39.7 is used in this study and was described and tested in Bayley et al. [23].

C.7 CODE AND DATA AVAILABILITY

The code to preprocess the observational data and to perform the fitting is published on the GitHub page: github.com/nilsnevertree/sdm-eurec4a. This repository also includes code to post-process the output data of the Cleo simulations and to produce the figures we use here. Post-processed simulation data can be provided upon reasonable request to the authors.

C.8 APPENDIX

C.8.1 Single-Droplet Mean-Field Approximation of Evaporation fraction

We consider an evaporating droplet of initial radius r_0 and mass m_0 falling in sub-saturated air ($q_v < q_{v,\text{sat}}$) from an initial height H . In classical evaporation theory [142, 208], the radius r of the droplet evolves as

$$\frac{dr^2}{dt} = -2D_* f_v \frac{\rho}{\rho_w} (q_{v,\text{sat}} - q_v), \quad (\text{C.2})$$

with

$$D_* = D_v \left/ \left[1 + \frac{\ell_v D_v \rho_{v,\text{sat}}}{k_a T} \left(\frac{\ell_v}{R_v T} - 1 \right) \right] \right. . \quad (\text{C.3})$$

D_v is the molecular diffusivity of water vapor in air, ℓ_v the enthalpy of vaporization, k_a the thermal conductivity of air, f_v the ventilation coefficient, and R_v the gas constant for water vapor. We neglect curvature and solute effects, which are negligible above $\sim 10\text{--}\mu\text{m}$. D_* is an effective diffusion coefficient that takes into account the thermal effects of evaporation [208].

We assume in this section that the droplet moves at its terminal velocity $U(r)$ at all times, which is reasonable on time scales longer than a few milliseconds. Numerical integration of the full equations of motion for a single droplet (not shown here) shows that the droplets are always within 0.01% of their terminal velocities, except near the very end of their trajectories if they evaporate completely. The height of the droplet can then be expressed as $dz = -U(r)dt$, so that Equation (C.2) becomes

$$\frac{dr^2}{dz} = 2 \frac{D_* f_v(r)}{U(r)} \frac{\rho}{\rho_w} (q_{v,\text{sat}} - q_v). \quad (\text{C.4})$$

We assume the vertical profiles of T and q_v to be uniform. Between cloud-base height H and the ground where $r(z = 0) = r_f$, Equation (C.4) then transforms into

$$\int_{r_0}^{r_f} \frac{U(r)}{f_v(r)} r dr = -D_* H \frac{\rho}{\rho_w} (q_{v,\text{sat}} - q_v) \quad (\text{C.5})$$

Exact closed-form solutions are difficult to find as they are set by the functional dependence of U and f_v with r . We can find an approximation in the low evaporation limit where $\xi \ll 1$ and the integral simplifies at first order, since U/f_v varies weakly:

$$\int_{r_0}^{r_f} \frac{U(r)}{f_v(r)} r dr \simeq \frac{U(r_0)}{f_v(r_0)} r_0 (r_f - r_0) \quad (\text{C.6})$$

so that

$$r_0 - r_f = \Delta r = D_* H \frac{f_v(r_0)}{U(r_0) r_0} \frac{\rho}{\rho_w} (q_{v,\text{sat}} - q_v). \quad (\text{C.7})$$

The evaporation fraction ξ reads

$$\xi = \frac{m_0 - m_f}{m_0} = \frac{r_0^3 - r_f^3}{r_0^3} = 1 - \left(1 - \frac{\Delta r}{r_0}\right)^3, \quad (\text{C.8})$$

which approximates, in the limit of $\Delta r \ll r_0$, to

$$\xi \simeq 3 \frac{\Delta r}{r_0} \simeq 3 D_* H \frac{f_v(r_0)}{U(r_0) r_0^2} \frac{\rho}{\rho_w} (q_{v,\text{sat}} - q_v). \quad (\text{C.9})$$

The dependence of the evaporation fraction with size is thus set by the size dependence of the ventilation coefficient and the terminal velocity. The terminal velocity can be reasonably approximated to be $U(r) = b\sqrt{r}$ between 500 μm and 2 mm. In the case of no ventilation effect ($f_v = 1$), this sets a power law dependence $\xi \propto r_0^{-5/2}$ in this limit. The ventilation coefficient varies as $f_v \propto N_{\text{Re}}^{1/2}$ with $N_{\text{Re}} \propto U(r_0) r_0$ the Reynolds number of the droplet [208]. This gives $\xi \propto r_0^{-7/4}$ when ventilation is taken into account.

c.8.2 Figures

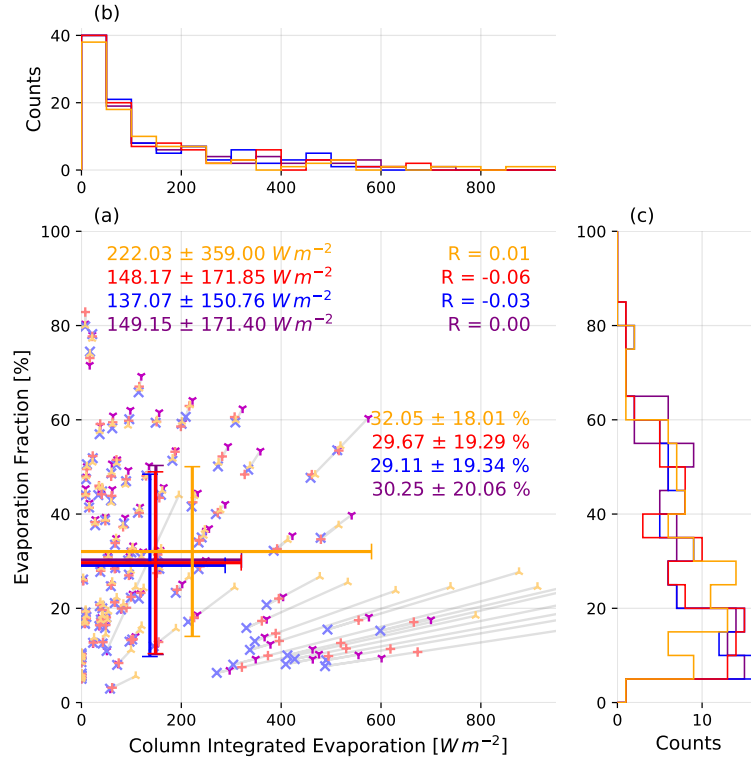


Figure C.11: (a) Evaporation fraction against column integrated evaporative cooling for individual clouds and all microphysics setups. Colors and markers as in Figure C.10 (EvapOnly in purple, EvapCoal in blue, EvapCoalBuRe-few in red and EvapCoalBuRe-many in orange). Simulations for the same cloud are connected by thin grey lines. Big error bars denote inter-cloud mean and standard deviation, both values denoted at top left and bottom right (Top to bottom: EvapCoalBuRe-many, EvapCoalBuRe-few, EvapCoal and EvapOnly). (b) Histogram of evaporation fraction for all microphysics setups. (c) Histogram of column integrated evaporative cooling for all microphysics setups.

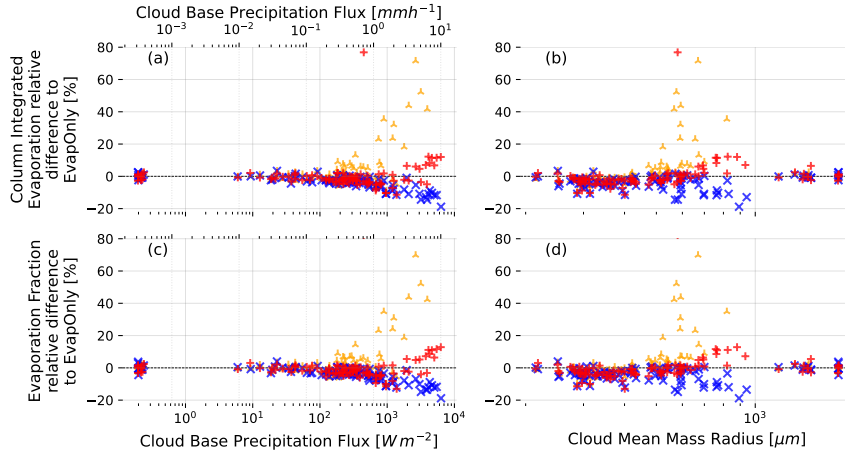


Figure C.12: Relative differences between EvapCoal and EvapCoalBuRe-few against EvapOnly of column integrated evaporative cooling (a,b) and evaporation fraction (c,d) compared to cloud-base precipitation flux (a,c) and cloud mean mass radius (b,d).

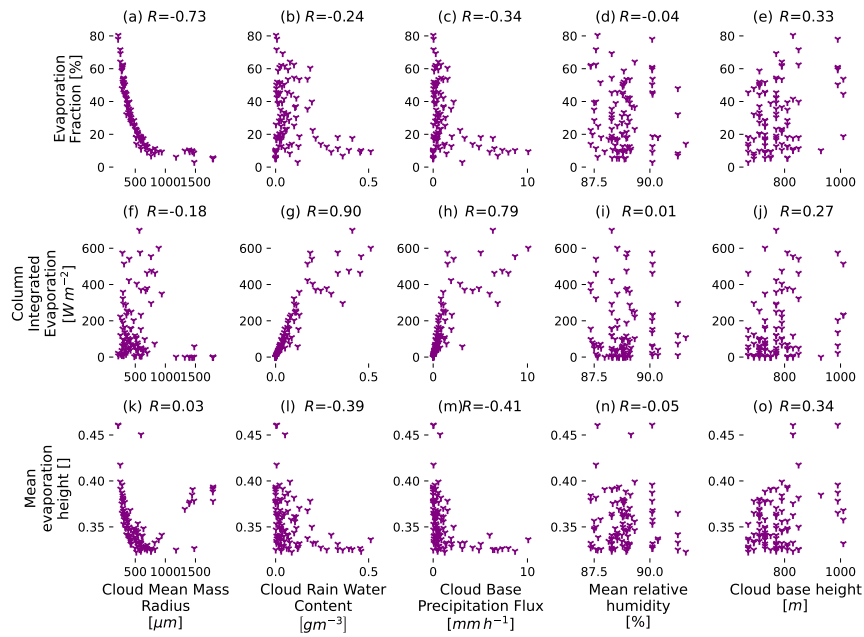


Figure C.13: Scatter of (a-e) evaporation fraction, (f-j) column integrated evaporative cooling and (k-o) mean evaporation height against various cloud and environmental variables.

C.9 COMPETING INTERESTS

CJAB declares that none of the authors have any competing interests.

C.10 ACKNOWLEDGEMENTS

Many thanks to the members of the rain evaporation club for our productive discussions. Thanks also to Shin-ichiro Shima (University of Hyogo, Japan) for his advice on the model setup and constructive feedback on the initial version of the manuscript. C. J.A. Bayley further expresses their appreciation for the work of the developers of the free and open-source software which underlies Cleo, especially from the developers of Git, GitHub, Python, the C++ standard libraries, and, above all, Kokkos.

R. Vogel acknowledges support from an ERC starting grant (ROTOЯ, grant no. 101116282). N. Niebaum, F. Poydenot and R. Vogel have further received funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2037 "Climate, Climatic Change, and Society" (project number 390683824)

A. K. Naumann has received funding which supported this work from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2037 "Climate, Climatic Change, and Society" (project number 390683824)

We thank the Deutsche Klimarechenzentrum (DKRZ) and the Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG) for the use of their computer facilities and software to conduct our work.

C.11 LARGE FIGURES

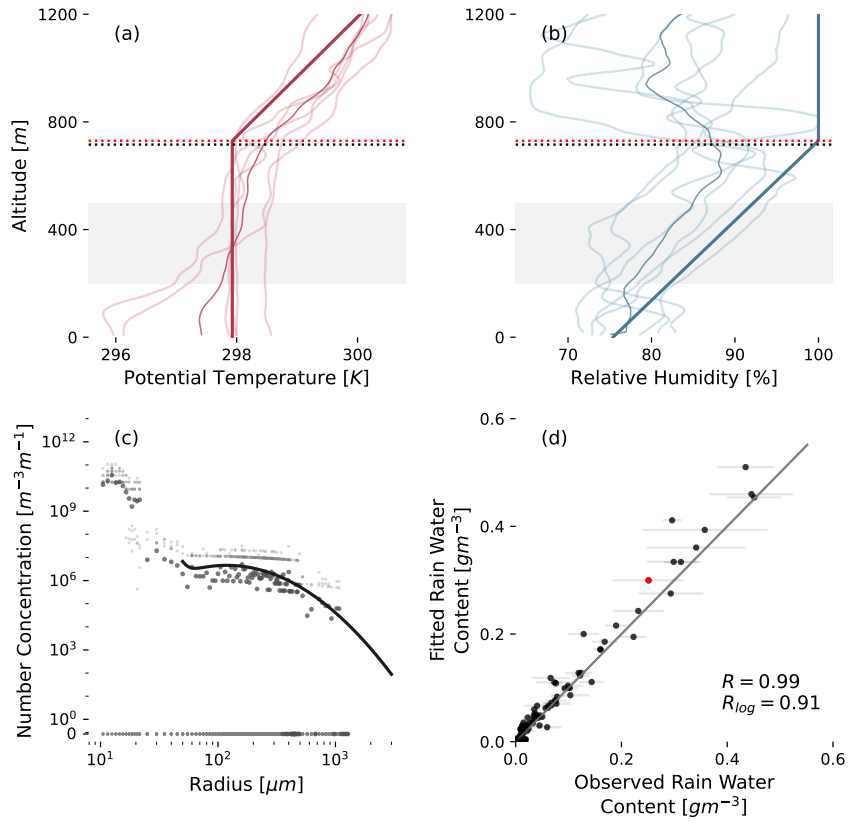


Figure C.14: Example of DSD and thermodynamic fits for a randomly selected cloud (a,b,c), typical of the entire cloud population we simulate. (a, b) The thermodynamic fits (thick dark line) overlaid on the 6 individual dropsonde measurements (thin light line) and their mean (thin dark line); (a) shows potential temperature (red); (b) shows relative humidity (blue). Data within the grey shaded altitude range from 200 – 500 m was used for the linear fits. The horizontal dashed black lines denote the flight altitude of the ATR aircraft, and the dashed red lines denote the altitude of 100% relative humidity given by the fit in panel (b). (c) The applied bimodal fit (solid black line) overlaid on the 23 s of ATR measurements of the number concentration of droplets in each radius bin (small light grey dots) and their temporal mean (large dark grey dots). Note that the y-axis changes from logarithmic to linear scale below 10^{-1} to show 0 values. (d) Comparison of the RWC of the fitted DSDs for each cloud used by the rainshaft model against the RWC from the observational cloud composite dataset as black dots. The error bars denote the standard error of the mean of the observations. The red dot indicates the example cloud used for panel (a,b,c).

	Terminal velocity	Evaporation / Condensation	Collisional Coalescence	Collisional Coalescence, Breakup & Rebound
Parametrisation	R1993; (see also B2025b)	explicit Köhler theory as in B2025b	L1974 hydrodynamic kernel as formulated by S2002	L1974 hydrodynamic kernel with coalescence / breakup / rebound efficiencies from Sch2010, Str2010; (see also B2025b)
EvapOnly	Yes	Yes	No	No
EvapCoal	Yes	Yes	Yes	No
EvapCoalBuRe-few	Yes	Yes	Yes	Yes (5)
EvapCoalBuRe-many	Yes	Yes	Yes	Yes (50)

Table C.1: Model setups used in this study and the parametrisations for each microphysical process. For collisional breakup, we prescribe a fixed number of fragments produced per breakup event, given in brackets. Acronyms in the table: R1993 = Rogers et al. [217], B2025b = Bayley et al. [23], L1974 = Long [159], S2002 = Simmel et al. [245], Sch2010 = Schlottkke et al. [230], and Str2010 = Straub et al. [266].

SDM AS A NUMERICAL BENCHMARK FOR WARM-RAIN PROCESSES?

The work in this appendix is currently in preparation for *Tellus*.

Authors: Clara J.A. Bayley^{1,2}, Emma Ware³, Sylwester Arabas⁴

¹Max-Planck-Institut für Meteorologie, Hamburg, Germany

²International Max Planck Research School on Earth System Modelling, Hamburg, Germany

³Department of Land, Air, and Water Resources, University of California Davis, Davis, CA, USA

⁴Faculty of Physics and Applied Computer Science, AGH University of Krakow, Poland

Author contributions: CJAB conceptualised the project and together with EW and SA developed the idea into the eventual structure of the study and interpreted the results. CJAB setup Cleo, ran both Cleo and PySDM, plotted the results and wrote the manuscript. EW implemented the alpha-sampling technique in PySDM, which Cleo also uses, and helped with running PySDM and plotting its results. EW and SA also setup PySDM and contributed their feedback to the manuscript.

D.1 ABSTRACT

Can Lagrangian particle models for cloud microphysics agree? For many reasons, the superdroplet method (SDM) appears to be a promising Lagrangian model for cloud microphysics which can eliminate or at least alleviate many of the sources of uncertainty which stem from conventional, bulk and bin, microphysics models. However, the recent 1-D Kinematic Driver-Aerosol (KiD-A) microphysics model inter-comparison (Hill et al. 2023, *J. Atmos. Sci.*, vol. 80) seemed to show the contrary, suggesting that model-related uncertainties in SDM cause large differences in precipitation, even in highly idealised setups and therefore cannot yet be used as a numerical benchmark for warm-rain microphysics modelling. In this study we revisit these conclusions, using the 1-D KiD framework with two SDMs, Cleo (v0.64.0) and PySDM (pre-v3), to test how sensitive they are to model-related uncertainties and whether or not two SDMs can agree. We explore uncertainty caused by superdroplet initialisation using the “alpha-sampling” technique, as well as differences in collision efficiencies and the number of superdroplets per grid-box. We find these SDM-related uncertainties cannot explain large differences in modelled precipitation and our SDMs can agree with one-another better than previously suggested. We also find that SDM in the KiD framework is highly sensitive to relatively small changes in the thermodynamics. We therefore provide the microphysics modelling community with an unambiguous thermodynamic setup for the KiD as well as a benchmark for warm-rain using SDM with a large enough number of superdroplets — one which tends towards the expected solutions from droplet-by-droplet simulations. This work supports the use of SDMs to bring consensus to microphysics modelling and to develop more affordable microphysics schemes, such as bulk models or neural networks. It also highlights the value of thorough model inter-comparison for microphysics model development.

D.2 INTRODUCTION

For decades uncertainty in warm-cloud microphysics has plagued both Large Eddy Simulations (LES) and global climate models (GCMs) [e.g. 2, 3, 37, 127, 130, 214, 289], and the advent of Global Storm Resolving Models (GSRMs) has further accentuated the need to understand and eliminate sources of microphysics uncertainty [e.g. 140, 176, 185, 262, 268, 272]. Uncertainty in microphysics is problematic because it allows differences between microphysics models which have serious impacts on the macro-scale. Spatial and temporal patterns of precipitation from warm shallow clouds vary significantly amongst LES using different microphysics schemes [2, 93, 213, 289], and on the global scale even changes within specific microphysics schemes can have large repercussions for tropical precipitation, sources of climate variability, and the Earth’s energy balance [17, 185, 196, 268, 272]. The high sensitivity of both regional and global models to differences in warm-cloud microphysics makes precipitation extremely challenging to predict, all the more so in response to aerosol perturbations and global warming. It is therefore essential to more tightly constrain the causes of these differences, the uncertainties in warm-cloud microphysics.

The sources of uncertainty in cloud microphysics can be broadly divided into two categories: knowledge and inherent model uncertainties. The former

relates to gaps in our physical understanding of cloud microphysics, for example how much droplet collision rates are increased by turbulent enhancement of droplet clustering or relative fall velocities. Inherent model uncertainties on the other hand, arise from *the way* we model the physics we already understand. That is, for a given physical description, all models are forced to make modelling choices which define how that physical description is represented. If these choices introduce poorly constrained parameters which the model outcome is highly sensitive to, then we introduce inherent model uncertainty into our simulations.

Conventional, bulk and bin, microphysics models give rise to various large inherent model uncertainties. Bulk schemes, also known as moment schemes, are useful for their comparatively low computational cost but they are forced to make problematic and somewhat arbitrary modelling assumptions to define different hydrometeor categories, and also the function form of the size distribution for each category (in the warm-phase for “cloud” vs. “rain” droplets) [65, 84, 90, 104, 106, 108–111, 130, 188]. Additional assumptions are then required to parametrise microphysical processes, which further ties bulk schemes to unobservable, and therefore difficult to constrain parameters [130, 207], so much so, that the names of real microphysics processes are often replaced with fictional ones, for example the process of “collisions between water droplets” is redefined as three distinct parametrisations called “self-collection”, “auto-conversion” and “accretion”. The sensitivity of macrophysical outcomes to the choices behind hydrometeor categories, size distributions, and microphysics parametrisations is high, and thus bulk schemes suffer from large inherent model uncertainty. Bin schemes eradicate the assumption of a function form for each hydrometeor size distribution, and in the warm-phase they do not need to categorise “cloud” and “rain” water separately. However, numerical diffusion in both spectral (size) and spatial space is inherent to bin schemes, as is their “curse of dimensionality”, and the mean-field approximation of the Smoluchowski equation used to describe droplet collisions [1, 56, 79, 84, 100, 117, 145, 146, 171, 181, 182, 254]. Different bin schemes may employ different methods for overcoming these limitations, especially with regard to numerical diffusion and their depiction of aerosols, and they may also use substantially different parametrisations for the same microphysical process, for example to depict activation and condensation/evaporation. As a result, bin schemes introduce sources of inherent model uncertainty, showing as large disagreement as bulk schemes when it comes to modelling precipitation [94, 103, 289].

The superdroplet method [SDM; also known as the superdroplet model; 241] is a fundamentally different way of modelling cloud microphysics and sets the expectation it can reduce inherent model uncertainties compared to conventional bulk/bin microphysics schemes. SDM is a Lagrangian, particle-in-cell method for cloud microphysics. It represents the real particle population with simulated particles called “superdroplets” (also known as “superparticles”), whereby each superdroplet has the properties of individual real particles, but also an extra property called their “multiplicity” which states how many real particles each superdroplet represents. The difference between SDM and other particle-in-cell methods for cloud microphysics is the collision algorithm, which conserves the number of superdroplets and converges to the results of particle-by-particle simulations faster than others [286]. Given its Lagrangian approach, SDM makes no assumptions about hydrometeor categories, nor size distributions, and it is intrinsically non-

diffusive and not limited by the Smoluchowski equation [84, 182]. Furthermore, the computational expense of incorporating more droplet and aerosol properties scales linearly in SDM, meaning it does not suffer from the “curse of dimensionality” of bin schemes [84, 182, 241]. SDM therefore eliminates many of the major sources of inherent model uncertainty from conventional bulk/bin microphysics. Additionally, for all the microphysical processes in warm-clouds: activation, condensation/evaporation, collisions and droplet sedimentation, the representation in SDM more explicitly relates to our fundamental physical description than it does in bulk or bin schemes, for example Köhler theory for activation and condensation/evaporation is applied to individual droplets rather than parametrised for droplet size distributions (DSDs). Consequently, SDM reduces the scope for model differences arising from microphysics parametrisations which convolute actual physics with model parameters. One would therefore expect SDM to alleviate inherent model uncertainty compared to conventional bulk/bin microphysics models, but this is only true provided macrophysical outcomes do not exhibit high sensitivity to SDM’s own model parameters.

Despite the expectation that SDM alleviates inherent model uncertainty, comparisons of SDMs in highly idealised setups seem to show otherwise and may prevent us obtaining a satisfactory numerical benchmark for warm-cloud precipitation. Unterstrasser et al. [286] found that choices regarding superdroplet initialisation affect the accuracy and convergence of the SDM collision algorithm, especially when the right-hand tail of the distribution is undersampled. Numerous studies also show sensitivity to the number of superdroplets per grid-box, and although $\mathcal{O}(10)$ to $\mathcal{O}(100)$ superdroplets per grid-box is generally considered sufficient for the mean statistics of collisions and/or precipitation, the exact number varies depending on the setup and the method of superdroplet initialisation [177, 286, 287, 311]. The most concerning results comparing SDMs came from Hill et al. [94], who used the Kinematic Driver (KiD) framework to compare nine different microphysics models. They found larger spread when they modelled precipitation amongst the three Lagrangian microphysics models, including two SDMs, than amongst the three bin and three bulk schemes. This might be expected given the latter are the outcome of several decades of work including tuning, whereas the Lagrangian models originated in the late 2000s. However, they also suggested that difference between the SDMs could be attributed to the choice of initialisation, number of simulated particles per grid-box, and/or the choice of interpolation for the collision efficiency in the collision kernel. All three suggestions concern somewhat arbitrary model parameters in SDM and therefore suggest that SDM has larger inherent model uncertainties than conventional bulk/bin models, in direct conflict with the idea that SDM would alleviate them. Based on these results Hill et al. [94] concluded we do not yet have a satisfactory numerical benchmark for warm-cloud precipitation processes and thus we cannot agree on how microphysics models should behave, even in highly idealised setups in the absence of knowledge uncertainties. Given the gravity of this conclusion, both for the future use of SDMs and for the broader community developing microphysics models, it warrants further investigation.

In this study we use two SDMs to revisit the whether or not SDM can alleviate inherent model uncertainty and be used as a numerical benchmark for warm-cloud processes. We use the 1-D KiD framework similarly to Hill et al. [94] with the SDMs Cleo (v0.64.0), [22, 23], and PySDM (pre-v3), [21,

123], as the microphysics schemes. To explore uncertainty caused by superdroplet initialisation we use the “alpha-sampling” technique described by Matsushima et al. [165] and to explore any differences caused by collision efficiencies we setup the SDMs with extremely different parametrisations. To eliminate any uncertainty in the KiD related to the number of superdroplets per grid-box, we also show the convergence of the mean precipitation and its error with increasing number of superdroplets. In these ways we sought to address the concerns raised by Hill et al. [94]. During our investigation, we also discovered that the KiD is highly sensitive to seemingly innocuous changes to the thermodynamic profiles and therefore we also discuss this non-SDM-related sensitivity. Based on our results, we conclude that it is possible for relatively small differences in the prescribed thermodynamics of the KiD to result in large differences in precipitation modelled by SDM. We therefore provide the microphysics modelling community with an unambiguous thermodynamic setup for the 1-D KiD as well as a step closer to an agreed benchmark for the setup using SDM with a large enough number of superdroplets.

Section D.3 describes the models and setup of the investigation, including the 1-D KiD, the microphysics schemes and the alpha sampling technique. We discuss our results for the various sources of uncertainty in Section D.4 as well as our progress towards a numerical benchmark. In Section D.5 we conclude with a summary of our findings and our reflections on using KiD model comparisons as a method for model development.

D.3 KID SETUP

In this section we describe the KiD models and setups we studied, including a detailed explanation of the thermodynamic profiles. This explanation is partly to resolve previous ambiguities in the setup so that other modellers can more easily run the KiD for their microphysics schemes under the same conditions. We also describe the two SDMs employed in this study and the method we used to initialise superdroplets in both models.

D.3.1 *KiD Model*

For our study the KiD model consists of either microphysics scheme, Cleo or PySDM, coupled to the dynamical driver provided by PyMPDATA, [20]. The PyMPDATA-based reimplementations of the KiD from Shipway and Hill [244] was originally developed for, and applied in the studies of De Jong et al. [51] and Azimi et al. [16]. To couple PyMPDATA to Cleo, we use Cleo’s python bindings. In this way our KiD models, using either Cleo or PySDM, are coupled to the same driver and therefore the models have the same advection scheme and thermodynamic setups but different SDM implementations, analogously to Hill et al. [94].

PyMPDATA prescribes the thermodynamics of the 1-D column and an updraught velocity for the first 10 min of the simulation which it uses to advect only the vapour field. Cleo and PySDM advect the hydrometeors and enact microphysical processes as explained in section D.3.2, but only the changes to the vapour, including its indirect effect on the pressure and density, are fed back into the dynamical driver, any temperature change due to condensation/evaporation is neglected. The column extends to 3200 m and has a grid-spacing of 25 m vertically and a horizontal cross-sectional

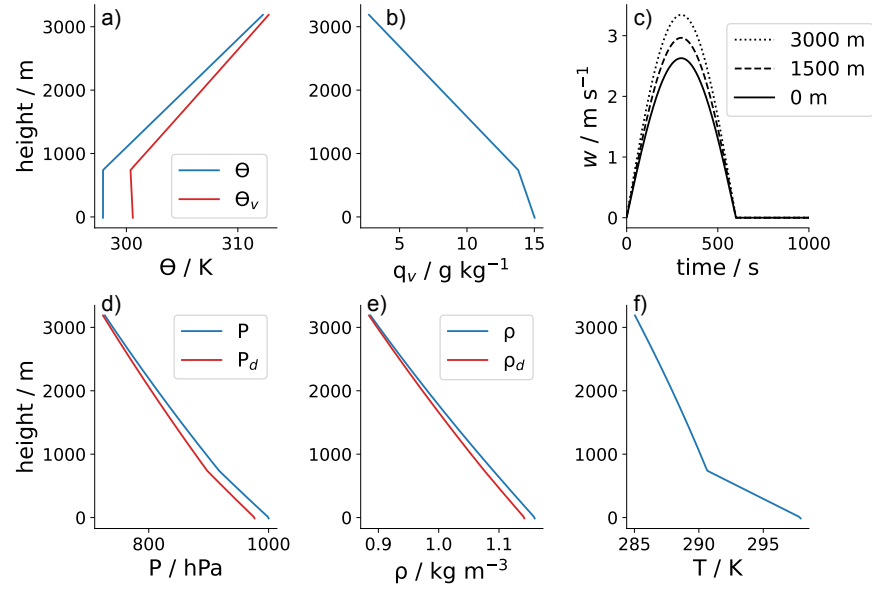


Figure D.1: The prescribed thermodynamics of our 1-D KiD model. *Top:* (a) The initial potential temperature and virtual potential temperature, θ and θ_v , respectively; (b) the initial water vapour mass mixing ratio q_v ; (c) the updraught velocity, w as a function of time at three different heights in the column. *Bottom:* (d) The initial total and dry pressure, P and P_d respectively; (e) the initial total and dry density, ρ and ρ_d respectively; (f) the initial temperature, T . The profiles of θ_v , P_d and ρ_d and T stay constant throughout the simulation whilst the water vapour is advected and responds to microphysics.

area of $1\ m^2$. This is equivalent to the specifications in [243] and Hill et al. [94], albeit with a different advection scheme and a different prescribed-flow, prescribed-temperature thermodynamic setup, as will be explained in the following.

We prescribe the thermodynamics similarly to the original 1-D KiD setup [93, 243, 244] and the KiD-A microphysics model inter-comparison [94], however we make some adjustments to clarify the setup and make it more consistent. The prescribed thermodynamics and updraught velocity for our setup are shown in Figure D.1. The initial vapour and potential temperature profiles are as described in the original 1-D KiD W2 and W3 setups [243, table 2], however this does not fully constrain the pressure, temperature, and density profiles in the column and these ambiguities make it hard for model comparisons to match the previous setups. Additionally the mass conservation of dry air was violated in the previous setups and the initial pressure and density profiles were not in hydrostatic equilibrium. We therefore chose to amend the original setup and to use hydrostatic pressure and density profiles with a surface pressure of 1000 hPa, also accounting for the mass of water vapour. The updraught velocity is similar to the original setup with $w_1 = 3\ m\ s^{-1}$ in Shipway and Hill [243, equation 6] except that we account for the dry air density profile in the column and advect the

momentum of dry air consistently rather than purely the updraught speed. Hence we use

$$\rho_d(z) w(z, t) = \begin{cases} \tilde{w}_1 \sin(\pi t / t_1), & \text{if } t < t_1, \\ 0.0, & \text{otherwise,} \end{cases} \quad (\text{D.1})$$

where $t_1 = 600 \text{ s}$ and $\tilde{w}_1 = 3 \text{ kg m}^{-2} \text{ s}^{-1}$.

D.3.1.1 *Alternative Thermodynamic Profiles*

During our investigation we also looked at the sensitivity of our results to three other thermodynamic profiles. We choose these thermodynamic profiles because whilst we were setting up our model comparison we realised these were all plausible variants of what can be interpreted from the previous 1-D KiD descriptions and, as will be discussed in Section D.4.3, we find that the cloud development and precipitation display high sensitivity to the relatively small differences between these thermodynamic profiles. The first variant ignores the effect of vapour on the hydrostatic profile. The second uses the Exner function as written in the source code of the original KiD paper [243]. The third profile uses the same profiles as the former, except that $\rho_d(z) = 1 \text{ kg m}^{-3}$ in Equation (D.1); in other words the density of air used for advection is kept constant throughout the column. This last setup is comparable to the setup in Hill et al. [94].

D.3.2 *Microphysics Schemes*

We use Cleo and PySDM with the microphysics summarised in Table D.1 in simulations with and without precipitation. In the no-precipitation simulations, Cleo and PySDM model condensation/evaporation and droplet motion which neglects their fall velocities. In the simulations with precipitation, the SDMs model condensation/evaporation, droplet motion including their fall velocities, and collision-coalescence. The particular algorithms used to solve for droplet motion differ between Cleo and PySDM, as do the parametrisations for droplet fall velocities. Whilst both SDMs explicitly solve the Köhler theory for condensation/evaporation, their parametrisations for the hygroscopicity of droplets and diffusion factors are different, and Cleo additionally accounts for ventilation effects which enhance the rate of condensation/evaporation of moving particles compared to stationary ones. Both SDMs use the original collision-coalescence algorithm according to Shima et al. [241] with their hydrodynamic collision kernel, but they differ due to their different parametrisations of the droplet fall velocities and collision-coalescence efficiencies, E_{coll} . PySDM assumes a uniform efficiency, $E_{\text{coll}} = 1.0$, whereas in this study Cleo uses either $E_{\text{coll}} = 1.0$, or the parametrisation with variable E_{coll} based on Long [159] according to Simmel and Wurzel [246]. Further details about the numerical methods for microphysics used in this study can be found for PySDM in Bartman et al. [21] and Jong et al. [123], and for Cleo in Bayley et al. [23].

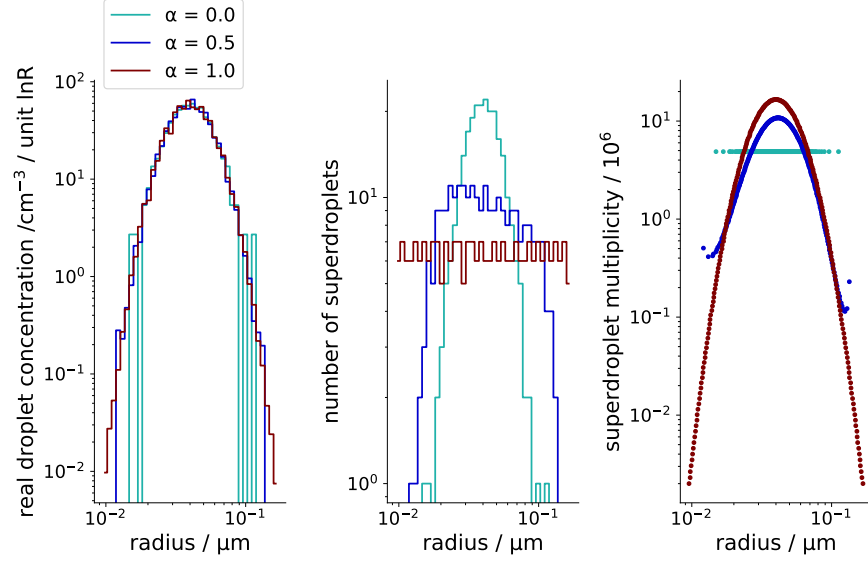


Figure D.2: An example of the superdroplet initialisation with alpha-sampling for three different values of α and 50 cm^{-3} total aerosol number concentration in the lowest 25 m of the domain. The real droplet size distribution (left) is reproduced from the superdroplet size distribution (middle) and their multiplicities (right). This example also highlights that lower values of α , towards constant-multiplicity sampling at $\alpha = 0.0$, less well sample the tails of the aerosol distribution.

D.3.3 Alpha Sampling for Superdroplet Initialisation

We use the initial aerosol conditions as described in Hill et al. [94]. Namely, we initialise dry NaCl aerosol with number concentrations, N_a , of 50, 150, and 300 cm^{-3} and a prescribed single-mode log-normal size distribution with a log-normal geometric mean radius of 0.04 mm and a log standard deviation of 1.4. We also weight the number concentration by the mass of dry air such that it decreases slightly higher in the column, although sensitivity tests showed the difference between that and a constant number concentration to be insignificant. During the simulation we treat the droplet boundary conditions slightly differently to Hill et al. [94] so that the bottom boundary of the domain is independent of the top and the superdroplet logic is improved. When $t < t_1$ and droplets are advected by the updraught, instead of droplets which leave the top of the domain being reused, they are removed from the simulation. New droplets are initialised according to the prescribed aerosol distribution and brought into equilibrium with the thermodynamic conditions of the bottom boundary of the domain before they are advected into the domain from the bottom boundary.

To initialise superdroplets we employ the alpha-sampling technique established by Matsushima et al. [165]. By default we sample the aerosol distribution with 256 superdroplets per-gridbox and we use the 1-D case of alpha-sampling, where the Wasserstein distance has a simpler form (see Appendix A therein), to sample superdroplet radii from the prescribed aerosol distribution. With this 1-D case, we define two end-points for the sampling method and a sampling-randomness. At one end-point the aerosol distri-

bution is sampled such that all superdroplets have equal multiplicities and the abundance of superdroplets with a certain radii follows the aerosol distribution. The second end-point is considered the opposite, whereby the multiplicities of the superdroplets establish the shape of the aerosol size distribution and superdroplet radii are on average equidistantly spaced in log-radius space. To move between these end-points, and thus change between different superdroplet sampling methods, we only need to change a single parameter, α . $\alpha = 0.0$ is the constant-multiplicity sampling and $\alpha = 1.0$ is the on average equidistantly spaced log-radius sampling, as demonstrated in Figure D.2. Mathematically, the chosen distribution sampled for superdroplet radii is that which minimises the α -weighted sum of the Wasserstein distances between that distribution and each end point distribution. The sampling-randomness determines the degree to which superdroplets are actually equidistantly spaced (in log-radius space when $\alpha = 1.0$) and this can be deterministic, pseudo-random or quasi-random. In the deterministic case, superdroplets are sampled at equidistant points along a distribution; in the pseudo-random case they are pseudo-randomly sampled freely across the entire distribution; and in the quasi-random case they are pseudo-randomly sampled from evenly spaced bins of the distribution.

The only other ways to change the superdroplet initialisation method would be to change the limits on the minimum and maximum radii we sample, or to sample radii not in log-space, for example in linear space. Since the minimum and maximum radii we sample represent the 0.00001 and 0.99999 percentiles of the droplet distribution we already comprehensively capture the initial aerosol population and so did not find it necessary to test the sensitivity of the KiD to changes in this sampling choice; and since previous literature has already clearly shown the improved performance of the SDM collision algorithm when using log-space sampling instead of linear-space sampling, because it better samples the right-hand tail of the distribution [e. g. 286], we did not test the sensitivity to changing this sampling choice either. For similar reasons, pseudo-random sampling of evenly spaced bins is usually considered superior to pseudo-random or deterministic sampling and so we also found it sufficient to use only the quasi-random sampling-randomness for this study. Therefore, we were left with one single parameter, α , which spans the maximum possible differences SDMs can have when they initialise superdroplet radii conventionally and quasi-randomly from (the 0.00001 to 0.99999 percentiles of) the chosen aerosol distribution.

D.4 RESULTS

In this section we present the results of the sensitivity tests we did to understand whether or not inherent model uncertainties in SDM related to the method for superdroplet initialisation, the E_{coll} parametrisation, or the number of superdroplets per grid-box are the cause of differences between precipitation patterns in the KiD framework. For all simulations with precipitation we show results for the mean and standard error of an ensemble of 10 members. The variance is caused by the pseudo-randomness of the SDM collision-coalescence algorithm and the quasi-randomness of sampling of the aerosol size distribution. For no-precipitation simulations, there is no detectable variance and so we only used one ensemble member.

D.4.1 Sensitivity to Initialisation and Collision Efficiencies

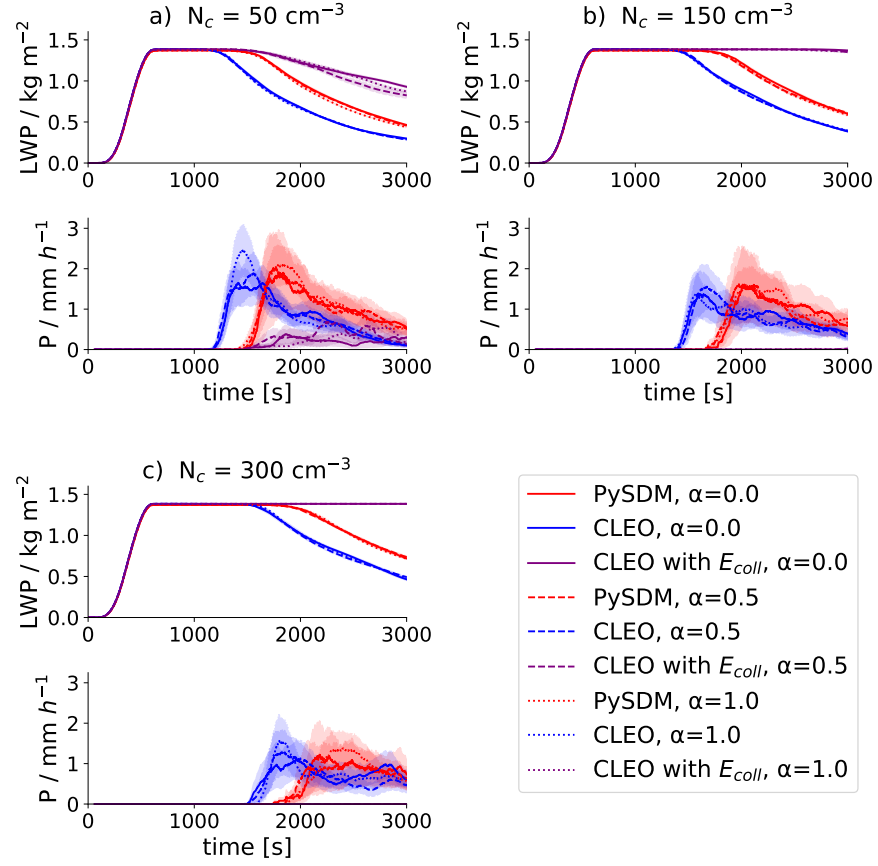


Figure D.3: The liquid water path (LWP) and precipitation as a function of time for the three different aerosol loadings, N_c , (a) 50 cm^{-3} , (b) 150 cm^{-3} , and (c) 300 cm^{-3} . The different lifestyles indicate different sampling methods via different values of α , the different colours indicate different microphysics models.

Differences in the sampling to initialise superdroplets cannot explain the spread amongst different SDMs in the KiD. Figure D.3 shows the liquid water path (LWP) and precipitation as a function of time for different microphysics setups in the KiD: PySDM with $E_{\text{coll}} = 1.0$, and Cleo with $E_{\text{coll}} = 1.0$ and Cleo with parametrised E_{coll} . None of the different microphysics setups completely agree with one another, and plausible reasons for this will be discussed in Section D.5.1. However, for each microphysics setup it is clear that within the ensemble error, changing the value of α has negligible effect on both the liquid water path (LWP) and precipitation. Of course, considerable effects could still be possible from choosing narrower upper and lower percentiles of the aerosol distribution to sample, or not using pseudo-random or log-space sampling. However, our results show that when sampling is conducted across a comprehensive range of the distribution in a conventional way, the sampling method in SDM is not a source of inherent model uncertainty.

Extremely large differences in collision efficiencies can explain some of the spread amongst different SDMs but it seems unpalatable. Comparing

Cleo with $E_{\text{coll}} = 1.0$ and Cleo with parametrised E_{coll} in Figure D.3, in the cleanest aerosol conditions it is possible to increase the maximum amount of precipitation by an order of magnitude and hasten precipitation onset by almost 10 min. In dirtier conditions it is possible to trigger precipitation with $E_{\text{coll}} = 1.0$ which would not occur otherwise. However the difference between these two parametrisations is extreme and using $E_{\text{coll}} = 1.0$ is unphysical. We hypothesise that the differences between realistic parametrisations of E_{coll} , for example between Hall [91] and Long [159], have a much smaller effect on precipitation, possibly even a negligible one, although further work is needed to test this hypothesis. However, even if small differences in collision efficiencies do cause some uncertainty in precipitation patterns from SDMs, they cannot explain it all. This is clear by comparing Cleo and PySDM both with $E_{\text{coll}} = 1.0$, which shows that even with identical E_{coll} precipitation onset can differ by almost 10 min (co-incidentally the same difference as between Cleo with uniform and parametrised E_{coll}). We draw from this that there must be other differences between Cleo and PySDM which are responsible for the remaining discrepancies in their KiD simulations and perhaps also the SDMs in Hill et al. [94]. These differences are unrelated to superdroplet initialisation and E_{coll} , and possibly more significant than any subtle differences in realistic E_{coll} parametrisations or the interpolation of them. Some plausible candidates will be discussed in Section D.4.3 and D.5.1.

D.4.2 Sensitivity to the Number of Superdroplets

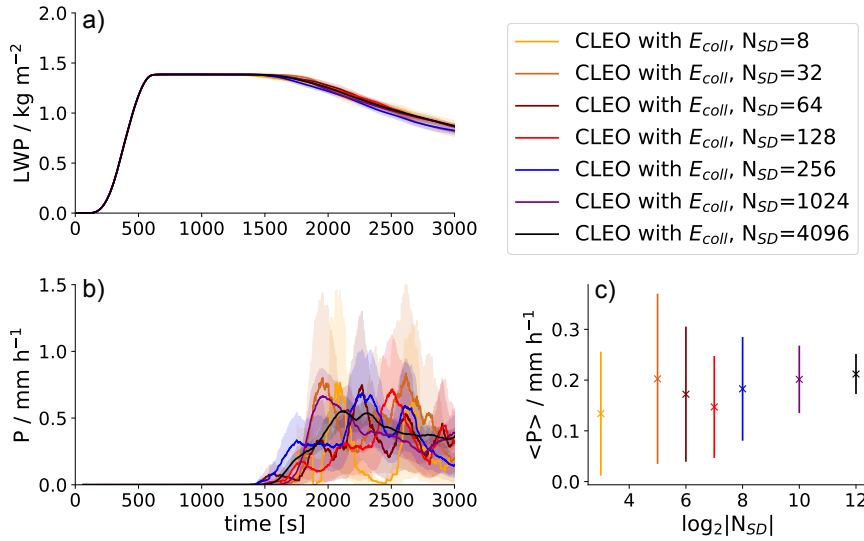


Figure D.4: (a) The liquid water path (LWP) and (b) precipitation as a function of time for the 50 cm^{-3} aerosol loading using Cleo with parametrised E_{coll} . (c) The mean and standard error of precipitation across the ensemble of simulations for a given number of superdroplets. The different colours from yellow to purple indicate increasing superdroplets per grid-box by the number of superdroplets sampled from the aerosol distribution.

Precipitation already shows reasonable convergence at $\mathcal{O}(10)$ superdroplets per grid-box. Figure D.4 shows the LWP varies negligibly between ensembles

with 8 to 4096 superdroplets sampling the initial aerosol distribution per grid-box. Whilst precipitation evolution shows much stochasticity, the timing of onset of precipitation and the mean precipitation amount are already well captured with 32 superdroplets per grid-box. These results are in agreement with Unterstrasser et al. [287] and Morrison et al. [177] which show that with mixing between collision-coalescence volumes the number of superdroplets needed for convergence is less than in box-model simulations and $\mathcal{O}(10)$ superdroplets per grid-box. Also as expected, the standard error decreases with increasing number of superdroplets as the variance in the collision-coalescence algorithm decreases. From these results we can conclude that with $\mathcal{O}(1000)$, this KiD setup with SDM is converged enough, and has small enough error, to make it practically useful as a numerical benchmark to compare with other microphysics schemes or SDMs.

D.4.3 Sensitivity to Thermodynamic Profiles

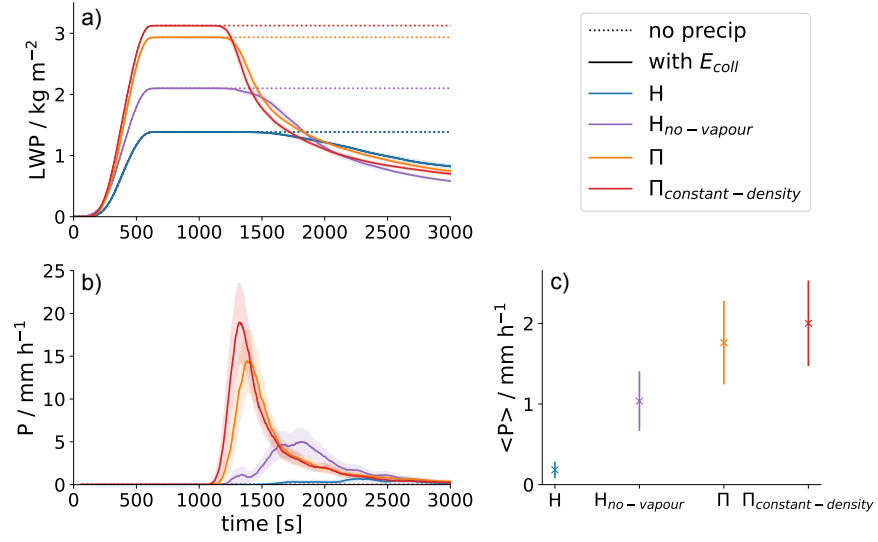


Figure D.5: (a) The liquid water path (LWP) and (b) precipitation as a function of time for the 50 cm^{-3} aerosol loading using Cleo with parametrised E_{coll} . Dotted lines show the no-precipitation simulations, solid lines with precipitation. (c) The mean and standard error of precipitation across the ensemble of simulations with precipitation for a given thermodynamics. The different colours indicate the different thermodynamic profiles, as described in Section D.3.1.1. H (blue) is the standard hydrostatic thermodynamics; $H_{\text{no-vapour}}$ (purple) neglects the contribution of water vapour pressure to the profile; Π (orange) uses the Exner function to calculate the thermodynamics; $\Pi_{\text{constant-density}}$ (red) also uses the Exner function but then sets the air density to be uniformly 1.0.

SDM is highly sensitive to seemingly innocuous changes to the thermodynamic profiles prescribed by the KiD. Whilst setting-up the comparison between Cleo and PySDM we noticed the description of the KiD thermodynamics contains ambiguities and so there are subtle differences in the assumptions one can make to derive temperature, pressure, density and

the updraught speed from the given potential temperature and initial water vapour profiles. Using four different derivations, which are the usual hydrostatic one and three variants (Section D.3.1.1), results in large differences in LWP and therefore precipitation as shown by Figure D.5. By not accounting for the initial water vapour density, a seemingly minor contribution to the hydrostatic profile, the LWP can nearly double and this results in almost a five-fold increase in mean precipitation. Using the Exner function to derive the thermodynamics results in even larger changes, and even the minor assumption of constant density for the updraught speed, raises the LWP still further, albeit less dramatically. This means that relatively minor differences in the KiD setup, show here by neglecting the effect of water vapour on density, and treating the updraught speed differently, can result in very large differences in the DSD evolution modelled by SDM, with consequent large differences in the macrophysical cloud evolution as well as precipitation. Such high sensitivity to the prescribed thermodynamics in the KiD is a plausible explanation for large differences between SDMs in Hill et al. [94], and it may explain the remaining differences between PySDM and Cleo in this study, as will be discussed in Section D.5.1.

D.5 CONCLUSION

In this study we have examined the sensitivity of LWP and precipitation to differences in the setup of 1-D KiD simulations with SDM to assess whether or not SDM can alleviate inherent model uncertainties and be used as a numerical benchmark for warm-rain microphysics modelling. We tested the sensitivity to three model-related parameters in SDM, suggested by Hill et al. [94] to be potentially significant sources of inherent model uncertainty. These were the method of superdroplet initialisation, the choice of collision efficiencies, E_{coll} , and the number of superdroplets per grid-box.

Using the alpha-sampling method from Matsushima et al. [165], it is clear that the sampling method used for superdroplet initialisation is a negligible source of uncertainty. Since changing the value of α has very little impact, our results for both Cleo and PySDM show no appreciable sensitivity to changing the sampling method all the way from constant-multiplicity to on average equidistantly log-spaced radii sampling. Such negligible sensitivity is provided that sampling is conducted across a comprehensive range of the aerosol distribution in a conventional way, in other words, with large enough maximum and minimum radii and with quasi-random log-space sampling rather than e.g. linear-space and/or pseudo-random or deterministic sampling.

Different collision efficiencies in SDM can explain some uncertainty in precipitation, but it is not a plausible explanation for the large differences reported by Hill et al. [94]. By comparing Cleo with $E_{\text{coll}} = 1.0$ and realistic parametrised E_{coll} , we can see that the amount of precipitation can change by an order of magnitude and delay precipitation onset by almost 10 min. However, we hypothesise that such large differences cannot occur if two realistic parametrised E_{coll} are used because they are much more similar than compared to $E_{\text{coll}} = 1.0$. Moreover, by comparing Cleo and PySDM when they both use $E_{\text{coll}} = 1.0$, we still observe different results which cannot be explained by differences in E_{coll} , with precipitation up-to 10 min later in PySDM. There must be other sensitivities at play.

When using SDM, minor differences in the prescribed thermodynamics of the KiD have an incommensurate impact on cloud development and precipitation and therefore could be a more significant contributor to differences between models than any inherent model uncertainties stemming from SDM. We have found that seemingly innocuous changes to the assumptions made to derive the prescribed thermodynamics for the KiD, for example excluding the effect of water vapour on density or deriving thermodynamics from the Exner function rather than hydrostatics, can drastically change the LWP modelled by SDM and consequently precipitation. The sensitivity of LWP to pressure/density assumptions has also been pointed out in previous discussions on SDM development [118], and such high sensitivity in the KiD is a plausible explanation for large differences between SDMs seen both in Hill et al. [94] and in our comparison of PySDM and Cleo with $E_{\text{coll}} = 1.0$. Due to this high sensitivity, it is important to precisely set the thermodynamic profiles for future model comparisons and to assist this we have carefully detailed our setup in Section D.3, including the thermodynamics in Section D.3.1.1.

The 1-D KiD setup is relatively insensitive to changing the number of superdroplets and we can therefore provide a useful numerical benchmark using SDM with $\mathcal{O}(1000)$ superdroplets per grid-box for the KiD. In agreement with previous studies, the mean precipitation from an ensemble of 1-D column SDM simulations already appears to be well converged at $\mathcal{O}(10)$ superdroplets per grid-box. By $\mathcal{O}(1000)$, the variance in the collision-coalescence algorithm is also small enough to make the precipitation modelled by SDM practically useful for comparison with other microphysics schemes. Of course, we can only provide a numerical benchmark which sets the “solution” for a given 1-D KiD setup, upon which all credible microphysics models should agree, once we find that using different SDMs produces consistent results, at least within what would be considered as a sufficiently small uncertainty range. At the time of writing, PySDM precipitates up-to 10 min later than Cleo, which makes the uncertainty in precipitation from these two SDMs far too large for a useful numerical benchmark. Further work is ongoing to understand and rectify this difference, as will be discussed in the following section.

Despite the fact we cannot yet provide a numerical benchmark, we have been able to rule out the supposed sources of inherent model uncertainty suggested by Hill et al. [94], and therefore our results reinvigorate support for the notion that SDM can reduce uncertainty in modelling warm-clouds by eliminating or at least alleviating inherent model uncertainty stemming from warm-cloud microphysics.

D.5.1 *Debugging and Model Development*

We have learnt both good and bad lessons from conducting this model comparison. Carefully analysing the differences between PySDM and Cleo led us to uncover many subtle but important factors in the KiD setup which were not previously documented. We uncovered and resolved ambiguities related to the thermodynamics (Section D.3.1.1), which, as already explained, are highly significant; and furthermore we uncovered and resolved ambiguities in the boundary conditions for the superdroplets (Section D.3.3). Along the way we also resolved a handful of bugs in our codes, both in our KiD setups and in our microphysics schemes, and implemented new features

in our models which their future use will benefit from. In these ways, the model comparison was an important and fruitful exercise. At the same time, we found that understanding model differences to the degree we can achieve consensus, even for just two SDMs in a highly idealised setup, often involves digging through source code and running perturbation experiments, both of which are time-consuming and often unrewarding tasks. We conclude that using the 1-D KiD framework to obtain model consensus is not just necessary for microphysics model development but also highly beneficial, however, it should be conducted with the understanding that achieving model agreement is surprisingly difficult.

Despite much de-bugging and tightly constraining the KiD setup, there is still almost a 10 min difference in the onset of precipitation from Cleo and PySDM. We are still working to understand this difference and have ruled-out many possible explanations including differences in the number concentration of aerosol, the hygroscopicity, the ventilation factor for condensation/evaporation, and almost all differences in the thermodynamic setup. Nevertheless, it appears that Cleo activates aerosol slightly more readily than PySDM and, counter-intuitively, precipitates earlier. We currently think that the difference in activation is not the cause of differences in the precipitation onset, and that differences in the probability of collisions, potentially due to the different calculation of relative fall velocities between droplets, is the cause. We also cannot yet definitively rule out remaining thermodynamic differences and we will never be absolutely certain there are no more bugs troubling our code. A thorough investigation of the difference in precipitation onset between our models is currently being conducted and will be reported in an updated version of the manuscript. As too will a discussion of the extent to which the remaining discrepancy and thermodynamic sensitivities we have already found also apply to bin and bulk schemes. We are nevertheless quietly confident we will manage to rectify the difference and thus use these SDMs in the 1-D KiD framework to provide a numerical benchmark for precipitation from warm-clouds.

D.6 ACKNOWLEDGEMENTS

CJAB thanks Bjorn Stevens for his dedicated supervision, discussion of results and constructive feedback on the manuscript. She also gratefully acknowledges the help developing Cleo from Tobias Kölling, Sergey Kosukhin and Lukas Kluft (Max Planck Institute for Meteorology, Germany; MPI-M).

We gratefully acknowledge the work of the developers and maintainers of numerous free/libre and open-source software packages, without which this study and the development of CLEO and PySDM would not have been possible. We also thank the Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG) from the information and communication services Cleo's development has benefited from, and finally we thank the Deutsche Klimarechenzentrum (DKRZ) for the computer facilities from project 1183 we used to conduct this work. Finally we thank the people at and funding from the MPI-M and Max-Planck-Gesellschaft (MPG) who/which made possible our participation in the workshop where the conceptualisation of this study then took place at Schloss Ringberg.

D.7 LARGE FIGURES

	Droplet Motion	Droplet Fall Velocities	Condensation and Evaporation	Collision-Coalescence
PySDM	Implicit Euler [11] with adaptive timestepping [123]	[89]	Predictor-corrector type [11] with adaptive timestepping [123]	[241] with $E_{\text{coll}} = 1.0$
Cleo	Heun's method [23, 78]	[246]	Implicit Euler with adaptive sub-stepping also including ventilation [23]	[241] with either $E_{\text{coll}} = 1.0$ or E_{coll} from [246]

Table D.1: Overview of the numerical methods used by the two SDMs used in this study.

BIBLIOGRAPHY

- ¹G. C. Abade, W. W. Grabowski and H. Pawlowska, ‘Broadening of Cloud Droplet Spectra through Eddy Hopping: Turbulent Entraining Parcel Simulations’, en, *Journal of the Atmospheric Sciences* **75**, 3365–3379 (2018).
- ²S. J. Abel and B. J. Shipway, ‘A comparison of cloud-resolving model simulations of trade wind cumulus with aircraft observations taken during RICO’, en, *Quarterly Journal of the Royal Meteorological Society* **133**, 781–794 (2007).
- ³A. S. Ackerman et al., ‘Large-Eddy Simulations of a Drizzling, Stratocumulus-Topped Marine Boundary Layer’, en, *Monthly Weather Review* **137**, 1083–1110 (2009).
- ⁴C. Acquistapace, R. Coulter, S. Crewell, A. Garcia-Benadi, R. Gierens, G. Labbri, A. Myagkov, N. Risse and J. H. Schween, ‘EUREC4A’s Maria S. Merian ship-based cloud and micro rain radar observations of clouds and precipitation’, English, *Earth System Science Data* **14**, 33–55 (2022).
- ⁵P. Alinaghi, M. Janssens and F. Jansson, *Warming from cold pools: A pathway for mesoscale organization to alter Earth’s radiation budget*, Version Number: 1, 2025.
- ⁶P. Alinaghi, F. Jansson, D. A. Blázquez and F. Glassmeier, ‘Cold pools mediate mesoscale adjustments of trade-cumulus fields to changes in cloud droplet number concentration’, en, *Atmospheric Chemistry and Physics* **25**, 6121–6139 (2025).
- ⁷P. Alinaghi, A. P. Siebesma, F. Jansson, M. Janssens and F. Glassmeier, ‘External Drivers and Mesoscale Self-Organization of Shallow Cold Pools in the Trade-Wind Regime’, en, *Journal of Advances in Modeling Earth Systems* **17**, e2024MS004540 (2025).
- ⁸N. Allwayin, D. J. Miller, K. K. Chandrakar, M. L. Larsen and R. A. Shaw, ‘Investigating Characteristic Droplet Size Distributions in Large Eddy Simulations of Stratocumulus Clouds’, en, *Geophysical Research Letters* **52**, e2025GL116021 (2025).
- ⁹M. Andrejczuk, W. W. Grabowski, J. Reisner and A. Gadian, ‘Cloud-aerosol interactions for boundary layer stratocumulus in the Lagrangian Cloud Model’, en, *Journal of Geophysical Research* **115**, D22214 (2010).
- ¹⁰M. Andrejczuk, J. M. Reisner, B. Henson, M. K. Dubey and C. A. Jeffery, ‘The potential impacts of pollution on a nondrizzling stratus deck: Does aerosol number matter more than type?’, en, *Journal of Geophysical Research* **113**, D19204 (2008).

- ¹¹S. Arabas, A. Jaruga, H. Pawlowska and W. W. Grabowski, 'Lib-cloudph++ 1.0: a single-moment bulk, double-moment bulk, and particle-based warm-rain microphysics library in C++', en, *Geoscientific Model Development* **8**, 1677–1707 (2015).
- ¹²S. Arabas and S.-i. Shima, 'Large-Eddy Simulations of Trade Wind Cumuli Using Particle-Based Microphysics with Monte Carlo Coalescence', en, *Journal of the Atmospheric Sciences* **70**, 2768–2777 (2013).
- ¹³S. Arabas and S.-i. Shima, 'On the CCN (de)activation nonlinearities', en, *Nonlinear Processes in Geophysics* **24**, 535–542 (2017).
- ¹⁴N. P. Arnold, W. M. Putman and S. R. Freitas, 'Impact of Resolution and Parameterized Convection on the Diurnal Cycle of Precipitation in a Global Nonhydrostatic Model', en, *Journal of the Meteorological Society of Japan. Ser. II* **98**, 1279–1304 (2020).
- ¹⁵O. Ayala, B. Rosa and L.-P. Wang, 'Effects of turbulence on the geometric collision rate of sedimenting droplets. Part 2. Theory and parameterization', *New Journal of Physics* **10**, 075016 (2008).
- ¹⁶S. Azimi, A. Jaruga, E. de Jong, S. Arabas and T. Schneider, 'Training Warm-Rain Bulk Microphysics Schemes Using Super-Droplet Simulations', *Journal of Advances in Modeling Earth Systems* **16**, [_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2023MS004028](https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2023MS004028), e2023MS004028 (2024).
- ¹⁷J. Bao and J. M. Windmiller, 'Impact of Microphysics on Tropical Precipitation Extremes in a Global Storm-Resolving Model', en, *Geophysical Research Letters* **48**, e2021GL094206 (2021).
- ¹⁸G. M. Barnes and M. Garstang, 'Subcloud Layer Energetics of Precipitating Convection', en, *Monthly Weather Review* **110**, 102–117 (1982).
- ¹⁹P. Bartman and S. Arabas, 'On the Design of Monte-Carlo Particle Coagulation Solver Interface: A CPU/GPU Super-Droplet Method Case Study with PySDM', en, in *Computational Science – ICCS 2021*, Vol. 12743, edited by M. Paszynski, D. Kranzlmüller, V. V. Krzhizhanovskaya, J. J. Dongarra and P. M. A. Sloot, Series Title: Lecture Notes in Computer Science (Springer International Publishing, Cham, 2021), pp. 16–30.
- ²⁰P. Bartman, J. Banaśkiewicz, S. Drenda, M. Manna, M. A. Olesik, P. Rozwoda, M. Sadowski and S. Arabas, 'PyMPDATA v1: Numba-accelerated implementation of MPDATA with examples in Python, Julia and Matlab', *Journal of Open Source Software* **7**, Publisher: The Open Journal, 3896 (2022).

- ²¹P. Bartman, O. Bulenok, K. Górski, A. Jaruga, G. Łazarski, M. A. Olesik, B. Piasecki, C. E. Singer, A. Talar and S. Arabas, ‘PySDM v1: particle-based cloud modeling package for warm-rain microphysics and aqueous chemistry’, *Journal of Open Source Software* **7**, 3219 (2022).
- ²²C. J. A. Bayley, T. Kölling, A. K. Naumann, R. Vogel and B. Stevens, ‘CLEO: The Fundamental Design for High Computational Performance of a New Superdroplet Model’, *EGUsphere* **2025**, 1–33 (2025).
- ²³C. J. A. Bayley, A. K. Naumann, F. Poydenot, R. Vogel, B. Stevens and S.-I. Shima, ‘CLEO: The Numerical Methods of a New Superdroplet Model including a Droplet Breakup Algorithm’, *EGUsphere* **2025**, 1–26 (2025).
- ²⁴C. J. A. Bayley, E. Ware and S. Arabas, ‘SDM as a Numerical Benchmark for Warm-Rain Processes? (in preparation)’, (2025).
- ²⁵C. Bayley, *CLEO: The Fundamental Design for High Computational Performance of a New Superdroplet Model [Dataset]*, Published: Edmond, <https://doi.org/10.17617/3.LNRKSJ>, 2025.
- ²⁶C. Bayley, *CLEO: The Numerical Methods of a New Superdroplet Model Including a Droplet Breakup Algorithm [Dataset]*, Published: Edmond, <https://doi.org/doi:10.17617/3.SDNoNX>, 2025.
- ²⁷C. J. A. Bayley, J. Behrens and G. Mania, *microphysics_testcases: github.com/yoctoyotta1024/microphysics_testcases*, (last accessed Jan. 2026).
- ²⁸C. J. A. Bayley, W. Loch, A. Devulapalli, L. Kluft, S. Kosukhin and T. Kölling, *CLEO: github.com/yoctoyotta1024/CLEO*, (last accessed Jan. 2026).
- ²⁹K. D. Beheng, K. Jellinghaus, W. Sander, N. Roth and B. Weigand, ‘Investigation of collision-induced breakup of raindrops by numerical simulations: First results’, in, *Geophysical Research Letters* **33**, 2005GL025519 (2006).
- ³⁰J. Bell Burnell, J. Farnham, D. Hill and A. Laking, *Beautiful Minds, Jocelyn Bell Burnell*, Type: Documentary Series, 2010.
- ³¹E. X. Berry, ‘A Mathematical Framework for Cloud Models’, in, *Journal of the Atmospheric Sciences* **26**, 109–111 (1969).
- ³²S. Bony et al., ‘EUREC4A observations from the SAFIRE ATR42 aircraft’, *Earth System Science Data* **14**, 2021–2064 (2022).
- ³³S. Bony, H. Schulz, J. Vial and B. Stevens, ‘Sugar, Gravel, Fish, and Flowers: Dependence of Mesoscale Patterns of Trade-Wind Clouds on Environmental Conditions’, in, *Geophysical Research Letters* **47**, e2019GL085988 (2020).
- ³⁴S. Bony et al., ‘Clouds, circulation and climate sensitivity’, in, *Nature Geoscience* **8**, 261–268 (2015).

- ³⁵S. Bony et al., 'EUREC4A: A Field Campaign to Elucidate the Couplings Between Clouds, Convection and Circulation', en, *Surveys in Geophysics* **38**, 1529–1568 (2017).
- ³⁶S. Brdar and A. Seifert, 'McSnow: A Monte-Carlo Particle Model for Riming and Aggregation of Ice Particles in a Multidimensional Microphysical Phase Space', en, *Journal of Advances in Modeling Earth Systems* **10**, 187–206 (2018).
- ³⁷C. S. Bretherton, 'Insights into low-latitude cloud feedbacks from high-resolution models', en, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **373**, 20140415 (2015).
- ³⁸V. Bringi, A. Seifert, W. Wu, M. Thurai, G.-J. Huang and C. Siewert, 'Hurricane Dorian Outer Rain Band Observations and 1D Particle Model Simulations: A Case Study', en, *Atmosphere* **11**, 879 (2020).
- ³⁹G. H. Bryan and H. Morrison, 'Sensitivity of a Simulated Squall Line to Horizontal Resolution and Parameterization of Microphysics', en, *Monthly Weather Review* **140**, 202–225 (2012).
- ⁴⁰C. J. A. Bayley* and N. Niebaum*, F. Poydenot, A. K. Naumann, M. Sarkar and R. Vogel, 'Constraining Rain Evaporation from Shallow-Clouds in the Trades using an Observation-Based Superdroplet Model', *EGUsphere* **2025**, 1–30 (2025).
- ⁴¹P. M. Caldwell et al., 'Convection-Permitting Simulations With the E3SM Global Atmosphere Model', en, *Journal of Advances in Modeling Earth Systems* **13**, e2021MS002544 (2021).
- ⁴²K. K. Chandrakar, W. W. Grabowski, H. Morrison and G. H. Bryan, 'Impact of entrainment-mixing and turbulent fluctuations on droplet size distributions in a cumulus cloud: An investigation using Lagrangian microphysics with a sub-grid-scale model', *Journal of the Atmospheric Sciences*, 10.1175/JAS-D-20-0281.1 (2021).
- ⁴³K. K. Chandrakar, H. Morrison, W. W. Grabowski and R. P. Lawson, 'Are turbulence effects on droplet collision-coalescence a key to understanding observed rain formation in clouds?', en, *Proceedings of the National Academy of Sciences* **121**, e2319664121 (2024).
- ⁴⁴K. K. Chandrakar, H. Morrison and R. A. Shaw, 'Lagrangian and Eulerian Supersaturation Statistics in Turbulent Cloudy Rayleigh-Bénard Convection: Applications for LES Subgrid Modeling', *Journal of the Atmospheric Sciences* **80**, Place: Boston MA, USA Publisher: American Meteorological Society, 2261–2285 (2023).
- ⁴⁵K. K. Chandrakar, H. Morrison and M. Witte, 'Evolution of Droplet Size Distributions During the Transition of an Ultraclean Stratocumulus Cloud System to Open Cell Structure: An LES Investigation Using Lagrangian Microphysics', en, *Geophysical Research Letters* **49**, e2022GL100511 (2022).

- ⁴⁶I. P. o. C. Change (IPCC), 'Evaluation of Climate Models', in *Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, 2014), pp. 741–866.
- ⁴⁷T. L. Clark and W. D. Hall, 'A Numerical Experiment on Stochastic Condensation Theory', en, *Journal of the Atmospheric Sciences* **36**, 470–483 (1979).
- ⁴⁸P. Coutris, *SAFIRE ATR42: PMA/Cloud composite dataset*, en, 2021.
- ⁴⁹C. Covey, P. J. Gleckler, C. Doutriaux, D. N. Williams, A. Dai, J. Fasullo, K. Trenberth and A. Berg, 'Metrics for the Diurnal Cycle of Precipitation: Toward Routine Benchmarks for Climate Models', en, *Journal of Climate* **29**, 4461–4471 (2016).
- ⁵⁰J. H. Curtis, N. Riemer and M. West, 'A single-column particle-resolved model for simulating the vertical distribution of aerosol mixing state: WRF-PartMC-MOSAIC-SCM v1.0', en, *Geoscientific Model Development* **10**, 4057–4079 (2017).
- ⁵¹E. De Jong, J. B. Mackay, O. Bulenok, A. Jaruga and S. Arabas, 'Breakups are complicated: an efficient representation of collisional breakup in the superdroplet method', en, *Geoscientific Model Development* **16**, 4193–4211 (2023).
- ⁵²V. K. Decyk and T. V. Singh, 'Particle-in-Cell algorithms for emerging computer architectures', en, *Computer Physics Communications* **185**, 708–719 (2014).
- ⁵³L. Denby, 'Discovering the Importance of Mesoscale Cloud Organization Through Unsupervised Classification', en, *Geophysical Research Letters* **47**, e2019GL085190 (2020).
- ⁵⁴B. J. Devenish et al., 'Droplet growth in warm turbulent clouds', en, *Quarterly Journal of the Royal Meteorological Society* **138**, 1401–1429 (2012).
- ⁵⁵W. Dong, M. Zhao, H. Guo, L. Harris, K.-y. Cheng, L. Zhou and V. Ramaswamy, 'Comparison of Global Mesoscale Convective System Simulations in a Global Storm-Resolving Model and a High-Resolution General Circulation Model', *Journal of Climate* **38**, 2339–2356 (2025).
- ⁵⁶P. Dziekan and H. Pawlowska, 'Stochastic coalescence in Lagrangian cloud microphysics', en, *Atmospheric Chemistry and Physics* **17**, 13509–13520 (2017).
- ⁵⁷P. Dziekan, M. Waruszewski and H. Pawlowska, 'University of Warsaw Lagrangian Cloud Model (UWLCM) 1.0: a modern large-eddy simulation tool for warm cloud modeling with Lagrangian microphysics', en, *Geoscientific Model Development* **12**, 2587–2606 (2019).

- ⁵⁸P. Dziekan and P. Zmijewski, 'University of Warsaw Lagrangian Cloud Model (UWLCM) 2.0: adaptation of a mixed Eulerian–Lagrangian numerical model for heterogeneous computing clusters', en, *Geoscientific Model Development* **15**, 4489–4501 (2022).
- ⁵⁹R. Eastman, I. L. McCoy, P. Zuidema, H. Schulz and P. Barrett, 'Increased Occurrence of Stratiform Clouds in the Caribbean Dry Season Since 1971 From Surface Observations', en, *Journal of Geophysical Research: Atmospheres* **130**, e2024JD042563 (2025).
- ⁶⁰H. C. Edwards, C. R. Trott and D. Sunderland, 'Kokkos: Enabling manycore performance portability through polymorphic memory access patterns', *Journal of Parallel and Distributed Computing* **74**, 3202–3216 (2014).
- ⁶¹F. Ewald, T. Kölling, A. Baumgartner, T. Zinner and B. Mayer, 'Design and characterization of specMACS, a multipurpose hyperspectral cloud and sky imager', en, *Atmospheric Measurement Techniques* **9**, 2015–2042 (2016).
- ⁶²N. M. Falk, A. L. Igel and M. R. Igel, 'The Relative Impact of Ice Fall Speeds and Microphysics Parameterization Complexity on Supercell Evolution', en, *Monthly Weather Review* **147**, 2403–2415 (2019).
- ⁶³Z. Feng, L. R. Leung, R. A. Houze, S. Hagos, J. Hardin, Q. Yang, B. Han and J. Fan, 'Structure and Evolution of Mesoscale Convective Systems: Sensitivity to Cloud Microphysics in Convection-Permitting Simulations Over the United States', en, *Journal of Advances in Modeling Earth Systems* **10**, 1470–1494 (2018).
- ⁶⁴S. Fiedler et al., 'Simulated Tropical Precipitation Assessed across Three Major Phases of the Coupled Model Intercomparison Project (CMIP)', en, *Monthly Weather Review* **148**, 3653–3680 (2020).
- ⁶⁵S. W. Freeman, A. L. Igel and S. C. Van Den Heever, 'Relative sensitivities of simulated rainfall to fixed shape parameters and collection efficiencies', en, *Quarterly Journal of the Royal Meteorological Society* **145**, 2181–2201 (2019).
- ⁶⁶B. Gasparini, R. Atlas, A. Voigt, M. Krämer and P. N. Blossey, *Tropical cirrus evolution in a km-scale model with improved ice microphysics*, Feb. 2025.
- ⁶⁷B. Gasparini, S. C. Sullivan, A. B. Sokol, B. Kärcher, E. Jensen and D. L. Hartmann, 'Opinion: Tropical cirrus – from micro-scale processes to climate-scale impacts', en, *Atmospheric Chemistry and Physics* **23**, 15413–15444 (2023).
- ⁶⁸G. George et al., 'JOANNE: Joint dropsonde Observations of the Atmosphere in tropical North atlantic meso-scale Environments', *Earth System Science Data* **13**, 5253–5272 (2021).
- ⁶⁹G. George, *JOANNE-Dataset: Joint dropsonde Observations of the Atmosphere in tropical North atlantic meso-scale Environments (v2.0.0)*, en, 2021.

- ⁷⁰G. George, B. Stevens, S. Bony, M. Klingebiel and R. Vogel, 'Observed impact of meso-scale vertical motion on cloudiness', *Journal of the Atmospheric Sciences*, **10.1175/JAS-D-20-0335.1** (2021).
- ⁷¹G. George, B. Stevens, S. Bony, R. Vogel and A. K. Naumann, 'Widespread shallow mesoscale circulations observed in the trades', en, *Nature Geoscience* **16**, 584–589 (2023).
- ⁷²A. Gettelman, H. Morrison, C. R. Terai and R. Wood, 'Microphysical process rates and global aerosol–cloud interactions', en, *Atmospheric Chemistry and Physics* **13**, 9855–9867 (2013).
- ⁷³D. T. Gillespie, 'The Stochastic Coalescence Model for Cloud Droplet Growth', en, *Journal of the Atmospheric Sciences* **29**, 1496–1510 (1972).
- ⁷⁴M. A. Giorgetta et al., 'ICON-A, the Atmosphere Component of the ICON Earth System Model: I. Model Description', en, *Journal of Advances in Modeling Earth Systems* **10**, 1613–1637 (2018).
- ⁷⁵A. M. Golovin, 'The Solution of the Coagulation Equation for Raindrops. Taking Condensation into Account', *Soviet Physics Doklady* **8**, 191 (1963).
- ⁷⁶W. W. Grabowski and L.-P. Wang, 'Diffusional and accretional growth of water drops in a rising adiabatic parcel: effects of the turbulent collision kernel', *Atmospheric Chemistry and Physics* **9**, 2335–2353 (2009).
- ⁷⁷W. W. Grabowski, 'Comparison of Eulerian Bin and Lagrangian Particle-Based Schemes in Simulations of Pi Chamber Dynamics and Microphysics', en, *JOURNAL OF THE ATMOSPHERIC SCIENCES* **77**, 15 (2020).
- ⁷⁸W. W. Grabowski, P. Dziekan and H. Pawlowska, 'Lagrangian condensation microphysics with Twomey CCN activation', en, *Geosci. Model Dev.*, 18 (2018).
- ⁷⁹W. W. Grabowski, 'Representation of Turbulent Mixing and Buoyancy Reversal in Bulk Cloud Models', en, *Journal of the Atmospheric Sciences* **64**, 3666–3680 (2007).
- ⁸⁰W. W. Grabowski, 'Extracting Microphysical Impacts in Large-Eddy Simulations of Shallow Convection', en, *Journal of the Atmospheric Sciences* **71**, 4493–4499 (2014).
- ⁸¹W. W. Grabowski, 'Comparison of Eulerian Bin and Lagrangian Particle-Based Microphysics in Simulations of Nonprecipitating Cumulus', *Journal of the Atmospheric Sciences* **77**, 3951–3970 (2020).
- ⁸²W. W. Grabowski and G. C. Abade, 'Broadening of Cloud Droplet Spectra through Eddy Hopping: Turbulent Adiabatic Parcel Simulations', en, *Journal of the Atmospheric Sciences* **74**, 1485–1493 (2017).

- ⁸³W. W. Grabowski, Y. Kim and S. S. Yum, 'CCN Activation and Droplet Growth in Pi Chamber Simulations with Lagrangian Particle-Based Microphysics', *Journal of the Atmospheric Sciences* **81**, 1201–1212 (2024).
- ⁸⁴W. W. Grabowski, H. Morrison, S.-I. Shima, G. C. Abade, P. Dziekan and H. Pawlowska, 'Modeling of Cloud Microphysics: Can We Do Better?', en, *Bulletin of the American Meteorological Society* **100**, 655–672 (2019).
- ⁸⁵W. W. Grabowski and H. Pawlowska, 'Technical note: Phase space depiction of cloud condensation nuclei activation and cloud droplet diffusional growth', en, *Atmospheric Chemistry and Physics* **25**, 5273–5285 (2025).
- ⁸⁶W. W. Grabowski and L. Thomas, 'Cloud droplet diffusional growth in homogeneous isotropic turbulence: bin microphysics versus Lagrangian super-droplet simulations', en, *Atmospheric Chemistry and Physics* **21**, 4059–4077 (2021).
- ⁸⁷G. Gramelsberger, J. Lenhard and W. Parker, 'Philosophical Perspectives on Earth System Modeling: Truth, Adequacy, and Understanding', en, *Journal of Advances in Modeling Earth Systems* **12**, e2019MS001720 (2020).
- ⁸⁸D. P. Grosvenor et al., 'Remote Sensing of Droplet Number Concentration in Warm Clouds: A Review of the Current State of Knowledge and Perspectives', en, *Reviews of Geophysics* **56**, 409–453 (2018).
- ⁸⁹R. Gunn and G. D. Kinzer, 'THE TERMINAL VELOCITY OF FALL FOR WATER DROPLETS IN STAGNANT AIR', en, *Journal of Meteorology* **6**, 243–248 (1949).
- ⁹⁰S. Hagos, L. Ruby Leung, C. Zhao, Z. Feng and K. Sakaguchi, 'How Do Microphysical Processes Influence Large-Scale Precipitation Variability and Extremes?', en, *Geophysical Research Letters* **45**, 1661–1667 (2018).
- ⁹¹W. D. Hall, 'A Detailed Microphysical Model Within a Two-Dimensional Dynamic Framework: Model Description and Preliminary Results', *Journal of Atmospheric Sciences* **37**, Place: Boston MA, USA Publisher: American Meteorological Society, 2486–2507 (1980).
- ⁹²M. Hanke, R. Redler, T. Holfeld and M. Yastremsky, 'YAC 1.2.0: new aspects for coupling software in Earth system modelling', *Geoscientific Model Development* **9**, 2755–2769 (2016).
- ⁹³A. A. Hill, B. J. Shipway and I. A. Boutle, 'How sensitive are aerosol-precipitation interactions to the warm rain representation?', en, *Journal of Advances in Modeling Earth Systems* **7**, 987–1004 (2015).
- ⁹⁴A. A. Hill et al., 'Toward a Numerical Benchmark for Warm Rain Processes', *Journal of the Atmospheric Sciences*, **10**. 1175/JAS-D-21-0275.1 (2023).

- ⁹⁵F. Hoffmann, 'The Effect of Spurious Cloud Edge Supersaturations in Lagrangian Cloud Models: An Analytical and Numerical Study', en, *Monthly Weather Review* **144**, 107–118 (2016).
- ⁹⁶F. Hoffmann, S. Raasch and Y. Noh, 'Entrainment of aerosols and their activation in a shallow cumulus cloud studied with a coupled LCM–LES approach', en, *Atmospheric Research* **156**, 43–57 (2015).
- ⁹⁷F. Hoffmann and G. Feingold, 'Cloud Microphysical Implications for Marine Cloud Brightening: The Importance of the Seeded Particle Size Distribution', *Journal of the Atmospheric Sciences* **78**, 3247–3262 (2021).
- ⁹⁸F. Hoffmann and G. Feingold, 'A Note on Aerosol Processing by Droplet Collision-Coalescence', en, *Geophysical Research Letters* **50**, e2023GL103716 (2023).
- ⁹⁹F. Hoffmann, Y. Noh and S. Raasch, 'The Route to Raindrop Formation in a Shallow Cumulus Cloud Simulated by a Lagrangian Cloud Model', en, *Journal of the Atmospheric Sciences* **74**, 2125–2142 (2017).
- ¹⁰⁰F. Hoffmann, T. Yamaguchi and G. Feingold, 'Inhomogeneous Mixing in Lagrangian Cloud Models: Effects on the Production of Precipitation Embryos', en, *Journal of the Atmospheric Sciences* **76**, 113–133 (2019).
- ¹⁰¹C. Hohenegger et al., 'ICON-Sapphire: simulating the components of the Earth system and their interactions at kilometer and subkilometer scales', *Geoscientific Model Development* **16**, 779–811 (2023).
- ¹⁰²W. E. Howell, 'THE GROWTH OF CLOUD DROPS IN UNIFORMLY COOLED AIR', en, *Journal of Meteorology* **6**, 134–149 (1949).
- ¹⁰³A. Z. Hu and A. L. Igel, 'A Bin and a Bulk Microphysics Scheme Can Be More Alike Than Two Bin Schemes', en, *Journal of Advances in Modeling Earth Systems* **15**, e2022MS003303 (2023).
- ¹⁰⁴A. Z. Hu and A. L. Igel, 'Bulk Microphysics Schemes May Perform Better With a Unified Cloud-Rain Category', en, *Journal of Advances in Modeling Earth Systems* **16**, e2023MS004068 (2024).
- ¹⁰⁵Z. Hu and R. C. Srivastava, 'Evolution of Raindrop Size Distribution by Coalescence, Breakup, and Evaporation: Theory and Observations', EN, *Journal of the Atmospheric Sciences* **52**, 1761–1783 (1995).
- ¹⁰⁶A. L. Igel, H. Morrison, S. P. Santos and M. Van Lier-Walqui, 'Limitations of Separate Cloud and Rain Categories in Parameterizing Collision-Coalescence for Bulk Microphysics Schemes', en, *Journal of Advances in Modeling Earth Systems* **14**, e2022MS003039 (2022).

- ¹⁰⁷A. L. Igel, 'Using an Arbitrary Moment Predictor to Investigate the Optimal Choice of Prognostic Moments in Bulk Cloud Microphysics Schemes', en, *Journal of Advances in Modeling Earth Systems* **11**, 4559–4575 (2019).
- ¹⁰⁸A. L. Igel, M. R. Igel and S. C. Van Den Heever, 'Make It a Double? Sobering Results from Simulations Using Single-Moment Microphysics Schemes', en, *Journal of the Atmospheric Sciences* **72**, 910–925 (2015).
- ¹⁰⁹A. L. Igel and S. C. Van Den Heever, 'The Importance of the Shape of Cloud Droplet Size Distributions in Shallow Cumulus Clouds. Part I: Bin Microphysics Simulations', en, *Journal of the Atmospheric Sciences* **74**, 249–258 (2017).
- ¹¹⁰A. L. Igel and S. C. Van Den Heever, 'The Importance of the Shape of Cloud Droplet Size Distributions in Shallow Cumulus Clouds. Part II: Bulk Microphysics Simulations', en, *Journal of the Atmospheric Sciences* **74**, 259–273 (2017).
- ¹¹¹A. L. Igel and S. C. Van Den Heever, 'The role of the gamma function shape parameter in determining differences between condensation rates in bin and bulk microphysics schemes', en, *Atmospheric Chemistry and Physics* **17**, 4599–4609 (2017).
- ¹¹²ISO, *ISO\slash IEC 14882:2020 Information technology — Programming languages — C++*, sixth (pub-ISO, pub-ISO:adr, Dec. 2020).
- ¹¹³M. Z. Jacobson, *Fundamentals of Atmospheric Modeling*, 2nd ed. (Cambridge University Press, 2005).
- ¹¹⁴M. Janssens, F. Jansson, P. Alinaghi, F. Glassmeier and A. P. Siebesma, 'Symmetry in Mesoscale Circulations Explains Weak Impact of Trade Cumulus Self-Organization on the Radiation Budget in Large-Eddy Simulations', en, *Geophysical Research Letters* **52**, e2024GL112288 (2025).
- ¹¹⁵M. Janssens, J. V.-G. De Arellano, C. C. Van Heerwaarden, S. R. De Roode, A. P. Siebesma and F. Glassmeier, 'Nonprecipitating Shallow Cumulus Convection Is Intrinsically Unstable to Length Scale Growth', *Journal of the Atmospheric Sciences* **80**, 849–870 (2023).
- ¹¹⁶M. Janssens, J. Vilà-Guerau De Arellano, M. Scheffer, C. Antonissen, A. P. Siebesma and F. Glassmeier, 'Cloud Patterns in the Trades Have Four Interpretable Dimensions', en, *Geophysical Research Letters* **48**, e2020GL091001 (2021).
- ¹¹⁷D. Jarecka, W. W. Grabowski, H. Morrison and H. Pawlowska, 'Homogeneity of the Subgrid-Scale Turbulent Mixing in Large-Eddy Simulation of Shallow Convection', en, *Journal of the Atmospheric Sciences* **70**, 2751–2767 (2013).

- ¹¹⁸A. Jaruga and H. Pawlowska, ‘Interactive comment on “libcloudph++ 1.1: aqueous phase chemistry extension of the Lagrangian cloud microphysics scheme” by Anna Jaruga and Hanna Pawlowska’, *Geoscientific Model Development Discussions*, **10**, 5194/gmd-2018-96-AC1 (2018).
- ¹¹⁹A. Jaruga and H. Pawlowska, ‘Libcloudph++ 2.0: aqueous-phase chemistry extension of the particle-based cloud microphysics scheme’, en, *Geoscientific Model Development* **11**, 3623–3645 (2018).
- ¹²⁰J. B. Jensen, S. Lee, P. B. Krummel, J. Katzfey and D. Gogoasa, ‘Precipitation in marine cumulus and stratocumulus.’, en, *Atmospheric Research* **54**, 117–155 (2000).
- ¹²¹B. Jian, J. Li, G. Wang, Y. Zhao, Y. Li, J. Wang, M. Zhang and J. Huang, ‘Evaluation of the CMIP6 marine subtropical stratocumulus cloud albedo and its controlling factors’, en, *Atmospheric Chemistry and Physics* **21**, 9809–9828 (2021).
- ¹²²T. Jokulsdottir and D. Archer, ‘A stochastic, Lagrangian model of sinking biogenic aggregates in the ocean (SLAMS 1.0): model formulation, validation and sensitivity’, en, *Geoscientific Model Development* **9**, 1455–1476 (2016).
- ¹²³E. K. de Jong et al., ‘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**, Publisher: The Open Journal, 4968 (2023).
- ¹²⁴F. Judt et al., ‘Tropical Cyclones in Global Storm-Resolving Models’, en, *Journal of the Meteorological Society of Japan. Ser. II* **99**, 579–602 (2021).
- ¹²⁵Y. Kajikawa, Y. Miyamoto, R. Yoshida, T. Yamaura, H. Yashiro and H. Tomita, ‘Resolution dependence of deep convections in a global simulation from over 10-kilometer to sub-kilometer grid spacing’, en, *Progress in Earth and Planetary Science* **3**, 16 (2016).
- ¹²⁶B. Kärcher and M. Corcos, ‘On the Lifetimes of Persistent Contrails and Contrail Cirrus’, en, *Journal of Geophysical Research: Atmospheres* **130**, e2025JD044488 (2025).
- ¹²⁷H. Kawai, S. Yukimoto, T. Koshiro, N. Oshima, T. Tanaka, H. Yoshimura and R. Nagasawa, ‘Significant improvement of cloud representation in the global climate model MRI-ESM2’, en, *Geoscientific Model Development* **12**, 2875–2897 (2019).
- ¹²⁸J. Kazil, R. Vogel, P. Blossey, S. Boeing, L. Denby, S. Ghazayel, T. Heus, R. Neggers, G. Raghunathan and P. Siebesma, ‘The Cold Pool Model Intercomparison Project (CP-MIP)’, in *European Geosciences Union General Assembly 2024 (EGU24)*, EGU General Assembly Conference Abstracts (Apr. 2024), p. 21182.

- ¹²⁹E. J. Kendon, N. M. Roberts, C. A. Senior and M. J. Roberts, 'Realism of Rainfall in a Very High-Resolution Regional Climate Model', en, *Journal of Climate* **25**, 5791–5806 (2012).
- ¹³⁰A. P. Khain et al., 'Representation of microphysical processes in cloud-resolving models: Spectral (bin) microphysics versus bulk parameterization', en, *Reviews of Geophysics* **53**, 247–322 (2015).
- ¹³¹J. M. King, C. D. Kummerow, S. C. Van Den Heever and M. R. Igel, 'Observed and Modeled Warm Rainfall Occurrence and Its Relationships with Cloud Macrophysical Properties', en, *Journal of the Atmospheric Sciences* **72**, 4075–4090 (2015).
- ¹³²G. D. Kinzer and R. Gunn, 'The Evaporation, Temperature and Thermal Relaxation-Time of Freely Falling Waterdrops.', *Journal of the Atmospheric Sciences* **8**, 71–83 (1951).
- ¹³³D. Klocke et al., 'Computing the full earth system at 1km resolution', in *Proceedings of the international conference for high performance computing, networking, storage and analysis*, SC '25 (2025), pp. 125–136.
- ¹³⁴H. Köhler, 'The nucleus in and the growth of hygroscopic droplets', *Trans. Faraday Soc.* **32**, Publisher: The Royal Society of Chemistry, 1152–1161 (1936).
- ¹³⁵A. B. Kostinski and R. A. Shaw, 'Fluctuations and Luck in Droplet Growth by Coalescence', en, *Bulletin of the American Meteorological Society* **86**, 235–244 (2005).
- ¹³⁶Kuan-Ting O, R. Wood and C. S. Bretherton, 'Ultraclean Layers and Optically Thin Clouds in the Stratocumulus-to-Cumulus Transition. Part II: Depletion of Cloud Droplets and Cloud Condensation Nuclei through Collision–Coalescence', en, *Journal of the Atmospheric Sciences* **75**, 1653–1673 (2018).
- ¹³⁷C. Kühnlein, W. Deconinck, R. Klein, S. Malardel, Z. P. Piotrowski, P. K. Smolarkiewicz, J. Szmelter and N. P. Wedi, 'FVM 1.0: a non-hydrostatic finite-volume dynamical core for the IFS', en, *Geoscientific Model Development* **12**, 651–676 (2019).
- ¹³⁸T. Kunihiro and K. Tsumura, 'Application of the renormalization-group method to the reduction of transport equations', *Journal of Physics A: Mathematical and General* **39**, 8089–8104 (2006).
- ¹³⁹I. La, W. W. Grabowski, Y. Kim, S. Kim and S. S. Yum, *Lagrangian Particle-Based Simulation of Aerosol-Dependent Vertical Variation of Cloud Microphysics in a Laboratory Convection Cloud Chamber*, Sept. 2025.
- ¹⁴⁰T. Lang, A. K. Naumann, S. A. Buehler, B. Stevens, H. Schmidt and F. Aemisegger, 'Sources of Uncertainty in Mid-Tropospheric Tropical Humidity in Global Storm-Resolving Simulations', en, *Journal of Advances in Modeling Earth Systems* **15**, e2022MS003443 (2023).

- ¹⁴¹R. Lange, 'ADPIC—A Three-Dimensional Particle-in-Cell Model for the Dispersal of Atmospheric Pollutants and its Comparison to Regional Tracer Studies', en, *Journal of Applied Meteorology* **17**, 320–329 (1978).
- ¹⁴²I. Langmuir, 'The Evaporation of Small Spheres', en, *Physical Review* **12**, 368–370 (1918).
- ¹⁴³K. M. Lau and H. T. Wu, 'Warm rain processes over tropical oceans and climate implications', en, *Geophysical Research Letters* **30**, 2003GL018567 (2003).
- ¹⁴⁴M. Leandro and P. Chuang, *ATOMIC aircraft microphysics: Size-resolved cloud and aerosol number concentrations taken from N43 aircraft in the North Atlantic Ocean, Barbados: Atlantic Tradewind Ocean-Atmosphere Mesoscale Interaction Campaign 2020-01-31 to 2020-02-10* (NCEI Accession 0232458), 2021.
- ¹⁴⁵H. Lee, A. M. Fridlind and A. S. Ackerman, 'An Evaluation of Size-Resolved Cloud Microphysics Scheme Numerics for Use with Radar Observations. Part I: Collision–Coalescence', en, *Journal of the Atmospheric Sciences* **76**, 247–263 (2019).
- ¹⁴⁶H. Lee, A. M. Fridlind and A. S. Ackerman, 'An evaluation of size-resolved cloud microphysics scheme numerics for use with radar observations Part II: Condensation and evaporation', *Journal of the Atmospheric Sciences*, **10.1175/JAS-D-20-0213.1** (2021).
- ¹⁴⁷T. Li, Y. Yu, X. Wang, X. Liu, Q. Mao and Y. Yuan, 'A review of aerosol-cloud interactions: Mechanisms, climate effects, and observation methods', en, *Atmospheric Research* **325**, 108267 (2025).
- ¹⁴⁸X.-Y. Li, A. Brandenburg, G. Svensson, N. E. L. Haugen, B. Mehlig and I. Rogachevskii, 'Effect of Turbulence on Collisional Growth of Cloud Droplets', en, *Journal of the Atmospheric Sciences* **75**, 3469–3487 (2018).
- ¹⁴⁹X.-Y. Li, A. Brandenburg, G. Svensson, N. E. L. Haugen, B. Mehlig and I. Rogachevskii, 'Condensational and Collisional Growth of Cloud Droplets in a Turbulent Environment', *Journal of the Atmospheric Sciences* **77**, 337–353 (2020).
- ¹⁵⁰X.-Y. Li, B. Mehlig, G. Svensson, A. Brandenburg and N. E. L. Haugen, 'Collision fluctuations of lucky droplets with superdroplets', en, *Journal of the Atmospheric Sciences* **79**, arXiv:1810.07475 [physics], 1821–1835 (2022).
- ¹⁵¹X.-Y. Li, G. Svensson, A. Brandenburg and N. E. L. Haugen, 'Cloud-droplet growth due to supersaturation fluctuations in stratiform clouds', en, *Atmospheric Chemistry and Physics* **19**, 639–648 (2019).
- ¹⁵²X.-Y. Li, A. Brandenburg, N. E. L. Haugen and G. Svensson, 'Eulerian and Lagrangian approaches to multidimensional condensation and collection', en, *Journal of Advances in Modeling Earth Systems* **9**, 1116–1137 (2017).

- ¹⁵³Z. Li, P. Zuidema, P. Zhu and H. Morrison, 'The Sensitivity of Simulated Shallow Cumulus Convection and Cold Pools to Microphysics', en, *Journal of the Atmospheric Sciences* **72**, 3340–3355 (2015).
- ¹⁵⁴J.-S. Lim, Y. Noh, H. Lee and F. Hoffmann, *The Critical Number and Size of Precipitation Embryos to Accelerate Warm Rain Initiation*, Aug. 2024.
- ¹⁵⁵J.-S. Lim and F. Hoffmann, 'Between Broadening and Narrowing: How Mixing Affects the Width of the Droplet Size Distribution', en, *Journal of Geophysical Research: Atmospheres* **128**, e2022JD037900 (2023).
- ¹⁵⁶J.-S. Lim and F. Hoffmann, 'Life Cycle Evolution of Mixing in Shallow Cumulus Clouds', en, *Journal of Geophysical Research: Atmospheres* **129**, e2023JD040393 (2024).
- ¹⁵⁷Y. Liu, M.-K. Yau, S.-i. Shima, C. Lu and S. Chen, 'Parameterization and Explicit Modeling of Cloud Microphysics: Approaches, Challenges, and Future Directions', en, *Advances in Atmospheric Sciences* **40**, 747–790 (2023).
- ¹⁵⁸G. Lloyd, A. Blyth, Z. Cui, T. Choularton, K. Bower, M. Gallagher, M. Flynn, N. Marsden, L. Denby and P. Gallimore, *The evolution of warm rain in trade-wind cumulus during EUREC⁴ A*, May 2024.
- ¹⁵⁹A. B. Long, *Solutions to the Droplet Collection Equation for Polynomial Kernels*, 1974.
- ¹⁶⁰E. N. Lorenz, 'The nature of the global circulation of the atmosphere: A present view, in The global circulation of the atmosphere', London, Royal Meteorological Society, 3–23 (1969).
- ¹⁶¹T. B. Low and R. List, 'Collision, Coalescence and Breakup of Raindrops. Part I: Experimentally Established Coalescence Efficiencies and Fragment Size Distributions in Breakup', en, *Journal of the Atmospheric Sciences* **39**, 1591–1606 (1982).
- ¹⁶²T. B. Low and R. List, 'Collision, Coalescence and Breakup of Raindrops. Part II: Parameterization of Fragment Size Distributions', en, *Journal of the Atmospheric Sciences* **39**, 1607–1619 (1982).
- ¹⁶³H.-Y. Ma, S. A. Klein, J. Lee, M.-S. Ahn, C. Tao and P. J. Gleckler, 'Superior Daily and Sub-Daily Precipitation Statistics for Intense and Long-Lived Storms in Global Storm-Resolving Models', en, *Geophysical Research Letters* **49**, e2021GL096759 (2022).
- ¹⁶⁴S. Matsugishi, T. Ohno and M. Satoh, *Differences in the cloud, precipitation, and convection representation between the global sub-km mesh simulation and km simulations*, Jan. 2025.
- ¹⁶⁵T. Matsushima, S. Nishizawa and S.-i. Shima, 'Overcoming computational challenges to realize meter- to submeter-scale resolution in cloud simulations using the super-droplet method', en, *Geoscientific Model Development* **16**, 6211–6245 (2023).

- ¹⁶⁶I. L. McCoy, 'Does It Matter to the Climate If Trade Cumulus Clouds Cluster?', en, *Geophysical Research Letters* **52**, e2025GL115570 (2025).
- ¹⁶⁷G. M. McFarquhar, 'A New Representation of Collision-Induced Breakup of Raindrops and Its Implications for the Shapes of Rain-drop Size Distributions', en, *Journal of the Atmospheric Sciences* **61**, 777–794 (2004).
- ¹⁶⁸D. L. McGuffin, D. D. Lucas, J. P. Morris, G. D. Spriggs and K. B. Knight, 'Super-Droplet Method to Simulate Lagrangian Microphysics of Nuclear Fallout in a Homogeneous Cloud', en, *Journal of Geophysical Research: Atmospheres* **127**, e2022JD036599 (2022).
- ¹⁶⁹J. D. McTaggart-Cowan and R. List, 'Collision and Breakup of Water Drops at Terminal Velocity', en, *Journal of the Atmospheric Sciences* **32**, 1401–1411 (1975).
- ¹⁷⁰M. Mech, E. Orlandi, S. Crewell, F. Ament, L. Hirsch, M. Hagen, G. Peters and B. Stevens, 'Hamp – the microwave package on the high altitude and long range research aircraft (halo)', *Atmospheric Measurement Techniques* **7**, 4539–4553 (2014).
- ¹⁷¹J. P. MELLADO, 'The evaporatively driven cloud-top mixing layer', *Journal of Fluid Mechanics* **660**, 5–36 (2010).
- ¹⁷²J. A. Milbrandt and M. K. Yau, 'A Multimoment Bulk Microphysics Parameterization. Part I: Analysis of the Role of the Spectral Shape Parameter', en, *Journal of the Atmospheric Sciences* **62**, 3051–3064 (2005).
- ¹⁷³J. A. Milbrandt and M. K. Yau, 'A Multimoment Bulk Microphysics Parameterization. Part II: A Proposed Three-Moment Closure and Scheme Description', en, *Journal of the Atmospheric Sciences* **62**, 3065–3081 (2005).
- ¹⁷⁴J. A. Milbrandt, M. K. Yau, J. Mailhot, S. Bélair and R. McTaggart-Cowan, 'Simulation of an Orographic Precipitation Event during IMPROVE-2. Part II: Sensitivity to the Number of Moments in the Bulk Microphysics Scheme', en, *Monthly Weather Review* **138**, 625–642 (2010).
- ¹⁷⁵A. A. Milne and E. H. Shepard, *Winnie-the-Pooh*, First (E.P. Dutton & Company, New York, 1926).
- ¹⁷⁶T. Miyakawa, M. Satoh, H. Miura, H. Tomita, H. Yashiro, A. T. Noda, Y. Yamada, C. Kodama, M. Kimoto and K. Yoneyama, 'Madden-Julian Oscillation prediction skill of a new-generation global model demonstrated using a supercomputer', en, *Nature Communications* **5**, 3769 (2014).

- ¹⁷⁷H. Morrison, K. K. Chandrakar, S.-I. Shima, P. Dziekan and W. W. Grabowski, 'Impacts of stochastic coalescence variability on warm rain initiation using Lagrangian microphysics in box and large-eddy simulations', *Journal of the Atmospheric Sciences*, **10.1175/JAS-D-23-0132.1** (2024).
- ¹⁷⁸H. Morrison and J. Milbrandt, 'Comparison of Two-Moment Bulk Microphysics Schemes in Idealized Supercell Thunderstorm Simulations', en, *Monthly Weather Review* **139**, 1103–1130 (2011).
- ¹⁷⁹H. Morrison and J. A. Milbrandt, 'Parameterization of Cloud Microphysics Based on the Prediction of Bulk Ice Particle Properties. Part I: Scheme Description and Idealized Tests', en, *Journal of the Atmospheric Sciences* **72**, 287–311 (2015).
- ¹⁸⁰H. Morrison, S. A. Tessendorf, K. Ikeda and G. Thompson, 'Sensitivity of a Simulated Midlatitude Squall Line to Parameterization of Raindrop Breakup', en, *Monthly Weather Review* **140**, 2437–2460 (2012).
- ¹⁸¹H. Morrison, M. Witte, G. H. Bryan, J. Y. Harrington and Z. J. Lebo, 'Broadening of Modeled Cloud Droplet Spectra Using Bin Microphysics in an Eulerian Spatial Domain', en, *Journal of the Atmospheric Sciences* **75**, 4005–4030 (2018).
- ¹⁸²H. Morrison et al., 'Confronting the Challenge of Modeling Cloud and Precipitation Microphysics', en, *Journal of Advances in Modeling Earth Systems* **12**, 10.1029/2019MS001689 (2020).
- ¹⁸³P. Narenpitak, J. Kazil, T. Yamaguchi, P. Quinn and G. Feingold, 'From Sugar to Flowers: A Transition of Shallow Cumulus Organization During ATOMIC', en, *Journal of Advances in Modeling Earth Systems* **13**, e2021MS002619 (2021).
- ¹⁸⁴A. K. Naumann and A. Seifert, 'Recirculation and growth of raindrops in simulated shallow cumulus', en, *Journal of Advances in Modeling Earth Systems* **8**, 520–537 (2016).
- ¹⁸⁵A. K. Naumann, M. Esch and B. Stevens, 'How the representation of microphysical processes affects tropical condensate in the global storm-resolving model ICON', en, *Atmospheric Chemistry and Physics* **25**, 6429–6444 (2025).
- ¹⁸⁶A. K. Naumann and C. Kiemle, 'The vertical structure and spatial variability of lower-tropospheric water vapor and clouds in the trades', en, *Atmospheric Chemistry and Physics* **20**, 6129–6145 (2020).
- ¹⁸⁷A. K. Naumann and A. Seifert, 'A Lagrangian drop model to study warm rain microphysical processes in shallow cumulus: A LAGRANGIAN DROP MODEL', en, *Journal of Advances in Modeling Earth Systems* **7**, 1136–1154 (2015).

- ¹⁸⁸A. K. Naumann and A. Seifert, 'Evolution of the Shape of the Rain-drop Size Distribution in Simulated Shallow Cumulus', en, *Journal of the Atmospheric Sciences* **73**, 2279–2297 (2016).
- ¹⁸⁹S. Nishizawa, H. Yashiro, Y. Sato, Y. Miyamoto and H. Tomita, 'Influence of grid aspect ratio on planetary boundary layer turbulence in large-eddy simulations', *Geoscientific Model Development* **8**, 3393–3419 (2015).
- ¹⁹⁰A. T. Noda, K. Oouchi, M. Satoh and H. Tomita, 'Quantitative Assessment of Diurnal Variation of Tropical Convection Simulated by a Global Nonhydrostatic Model without Cumulus Parameterization', en, *Journal of Climate* **25**, 5119–5134 (2012).
- ¹⁹¹Y. Noh, D. Oh, F. Hoffmann and S. Raasch, 'A Cloud Microphysics Parameterization for Shallow Cumulus Clouds Based on Lagrangian Cloud Model Simulations', en, *Journal of the Atmospheric Sciences* **75**, 4031–4047 (2018).
- ¹⁹²L. Nuijens and A. P. Siebesma, 'Boundary Layer Clouds and Convection over Subtropical Oceans in our Current and in a Warmer Climate', *Current Climate Change Reports* **5**, 80–94 (2019).
- ¹⁹³L. Nuijens, B. Stevens and A. P. Siebesma, 'The Environment of Precipitating Shallow Cumulus Convection', en, *Journal of the Atmospheric Sciences* **66**, 1962–1979 (2009).
- ¹⁹⁴D. Oh, Y. Noh and F. Hoffmann, 'Paths From Aerosol Particles to Activation and Cloud Droplets in Shallow Cumulus Clouds: The Roles of Entrainment and Supersaturation Fluctuations', en, *Journal of Geophysical Research: Atmospheres* **128**, e2022JD038450 (2023).
- ¹⁹⁵I. Orlanski, 'A Rational Subdivision of Scales for Atmospheric Processes', *Bulletin of the American Meteorological Society* **56**, Publisher: American Meteorological Society, 527–530 (1975).
- ¹⁹⁶S. Ortega, H. Segura, V. Mayta, R. Fievet, A. Peinado, J. Lee, M. Giorgetta and B. Stevens, 'The representation of Convectively Coupled Equatorial Waves in a Global Storm Resolving Model', *Journal of Advances in Modeling Earth Systems* *submitted* (2026).
- ¹⁹⁷R. K. T. Oware, Y. Zheng, P. Bogenschutz, Y. Zhang, H.-Y. Ma, S. Xie and C. Tao, 'Evaluating E3SM Global Storm-Resolving Model Simulations of Deep Convection: Insights From DP-SCREAM During TRACER', en, *Journal of Geophysical Research: Atmospheres* **130**, e2025JD044113 (2025).
- ¹⁹⁸T. Palmer and B. Stevens, 'The scientific challenge of understanding and estimating climate change', en, *Proceedings of the National Academy of Sciences* **116**, 24390–24395 (2019).
- ¹⁹⁹I. R. Paluch and D. H. Lenschow, 'Stratiform Cloud Formation in the Marine Boundary Layer', en, *Journal of the Atmospheric Sciences* **48**, 2141–2158 (1991).

- ²⁰⁰A. Peinado, 'Convergence of the simulated Tropical Convergence Zone and its response to climate change with resolution', en, *Art-work Size: 17761279 Medium: application/pdf Publisher: MPI für Meteorologie Version Number: 1, 17761279* (2025).
- ²⁰¹R. Pincus et al., 'Observations from the NOAA P-3 aircraft during ATOMIC', English, *Earth System Science Data* **13**, 3281–3296 (2021).
- ²⁰²Piotr Dziekan, Jørgen B. Jensen, Wojciech W. Grabowski and Hanna Pawlowska, 'Impact of Giant Sea Salt Aerosol Particles on Precipitation in Marine Cumuli and Stratocumuli: Lagrangian Cloud Model Simulations', *Journal of the Atmospheric Sciences* **78**, 4127–4142 (2021).
- ²⁰³C. Planche, F. Tridon, S. Banson, G. Thompson, M. Monier, A. Battaglia and W. Wobrock, 'On the Realism of the Rain Microphysics Representation of a Squall Line in the WRF Model. Part II: Sensitivity Studies on the Rain Drop Size Distributions', en, *Monthly Weather Review* **147**, 2811–2825 (2019).
- ²⁰⁴V. T. Pörtge, 'Understanding cloud droplet size distributions from multi-angle polarimetric observations' (May 2024).
- ²⁰⁵O. P. Prat, A. P. Barros and F. Y. Testik, 'On the Influence of Rain-drop Collision Outcomes on Equilibrium Drop Size Distributions', en, *Journal of the Atmospheric Sciences* **69**, 1534–1546 (2012).
- ²⁰⁶J. du Pré, S. J. Barbirolli, London Symphony Orchestra and E. Elgar, *Cello Concerto in E Minor, Op. 85: III. Adagio*, Recorded in Kingsway Hall, London; du Pré on cello, Barbirolli as conductor, Elgar composer, 1965.
- ²⁰⁷U. Proske, S. Ferrachat, S. Klampt, M. Abeling and U. Lohmann, 'Addressing Complexity in Global Aerosol Climate Model Cloud Microphysics', en, *Journal of Advances in Modeling Earth Systems* **15**, e2022MS003571 (2023).
- ²⁰⁸H. R. Pruppacher and J. D. Klett, *Microphysics of Clouds and Precipitation*, English, 1st ed. (D. Reidel Publishing Company, 1978).
- ²⁰⁹H. R. Pruppacher and R. Rasmussen, 'A Wind Tunnel Investigation of the Rate of Evaporation of Large Water Drops Falling at Terminal Velocity in Air', *Journal of Atmospheric Sciences* **36**, Place: Boston MA, USA Publisher: American Meteorological Society, 1255–1260 (1979).
- ²¹⁰W. M. Putman and M. Suarez, 'Cloud-system resolving simulations with the NASA Goddard Earth Observing System global atmospheric model (GEOS-5): GEOS5 CLOUD-SYSTEM RESOLVING SIMULATIONS', en, *Geophysical Research Letters* **38**, n/a–n/a (2011).
- ²¹¹J. Radtke and U. Hamburg, 'On the relationship between precipitation and spatial organization in the trades', en, (2023).

- ²¹²J. Radtke, A. K. Naumann, M. Hagen and F. Ament, 'The relationship between precipitation and its spatial pattern in the trades observed during EUREC4A', en, *Quarterly Journal of the Royal Meteorological Society* **148**, 1913–1928 (2022).
- ²¹³G. N. Raghunathan, P. Blossey, S. Boeing, L. Denby, S. Ghazayel, T. Heus, J. Kazil and R. Neggers, 'Flower-Type Organized Trade-Wind Cumulus: A Multi-Day Lagrangian Large Eddy Simulation Intercomparison Study', en, *Journal of Advances in Modeling Earth Systems* **17**, e2024MS004864 (2025).
- ²¹⁴D. Randall, M. Khairoutdinov, A. Arakawa and W. Grabowski, 'Breaking the Cloud Parameterization Deadlock', en, *Bulletin of the American Meteorological Society* **84**, 1547–1564 (2003).
- ²¹⁵T. Riechelmann, Y. Noh and S. Raasch, 'A new method for large-eddy simulations of clouds with Lagrangian droplets including the effects of turbulent collision', *New Journal of Physics* **14**, 065008 (2012).
- ²¹⁶N. Rochetin, C. Hohenegger, L. Touzé-Peiffer and N. Villefranque, 'A Physically Based Definition of Convectively Generated Density Currents: Detection and Characterization in Convection-Permitting Simulations', en, *Journal of Advances in Modeling Earth Systems* **13**, e2020MS002402 (2021).
- ²¹⁷R. R. Rogers, D. Baumgardner, S. A. Ethier, D. A. Carter and W. L. Ecklund, 'Comparison of Raindrop Size Distributions Measured by Radar Wind Profiler and by Airplane', *Journal of Applied Meteorology and Climatology* **32**, Place: Boston MA, USA Publisher: American Meteorological Society, 694–699 (1993).
- ²¹⁸Y. M. Rogers R.R., *A Short Course in Cloud Physics*, English, 3rd ed. (Butterworth-Heinemann, Jan. 1989).
- ²¹⁹W. B. Rossow, A. Mekonnen, C. Pearl and W. Goncalves, 'Tropical Precipitation Extremes', en, *Journal of Climate* **26**, 1457–1466 (2013).
- ²²⁰K. Sakaguchi, T. Suematsu, Z. Feng, H. Miura, B. E. Harrop, L. Ruby, W. C. Skamarock and M. Duda, 'Tuning Physics Parameters for Tropical Convective Systems Across Scales in Two Global Convection-Permitting Models', in AGU Fall Meeting Abstracts, Vol. 2022, AGU Fall Meeting Abstracts (Dec. 2022), A51I–03.
- ²²¹I. Sandu and B. Stevens, 'On the Factors Modulating the Stratocumulus to Cumulus Transitions', en, *Journal of the Atmospheric Sciences* **68**, 1865–1881 (2011).
- ²²²M. Sarkar, A. Bailey, P. Blossey, S. P. De Szoeki, D. Noone, E. Quiñones Meléndez, M. D. Leandro and P. Y. Chuang, 'Sub-cloud rain evaporation in the North Atlantic winter trade winds derived by pairing isotopic data with a bin-resolved microphysical model', en, *Atmospheric Chemistry and Physics* **23**, 12671–12690 (2023).

- ²²³M. Sarkar, R. Vogel and Y. Zheng, *Microphysics Dominates the Sub-Cloud Rain Evaporation in the North Atlantic Trades*, Jan. 2025.
- ²²⁴M. Sarkar, P. Zuidema, B. Albrecht, V. Ghate, J. Jensen, J. Mohrmann and R. Wood, 'Observations Pertaining to Precipitation within the Northeast Pacific Stratocumulus-to-Cumulus Transition', en, *Monthly Weather Review* **148**, 1251–1273 (2019).
- ²²⁵T. Sato, H. Miura, M. Satoh, Y. N. Takayabu and Y. Wang, 'Diurnal Cycle of Precipitation in the Tropics Simulated in a Global Cloud-Resolving Model', en, *Journal of Climate* **22**, 4809–4826 (2009).
- ²²⁶Y. Sato, S. Nishizawa, H. Yashiro et al., 'Impacts of cloud microphysics on trade wind cumulus: which cloud microphysics processes contribute to the diversity in a large eddy simulation?', *Progress in Earth and Planetary Science* **2**, 23 (2015).
- ²²⁷Y. Sato, S.-i. Shima and H. Tomita, 'A grid refinement study of trade wind cumuli simulated by a Lagrangian cloud microphysical model: the super-droplet method: Grid refinement study of cumuli using the super-droplet method', en, *Atmospheric Science Letters* **18**, 359–365 (2017).
- ²²⁸Y. Sato, S.-i. Shima and H. Tomita, 'Numerical Convergence of Shallow Convection Cloud Field Simulations: Comparison Between Double-Moment Eulerian and Particle-Based Lagrangian Microphysics Coupled to the Same Dynamical Core', en, *Journal of Advances in Modeling Earth Systems* **10**, 1495–1512 (2018).
- ²²⁹M. Satoh, B. Stevens, F. Judt, M. Khairoutdinov, S.-J. Lin, W. M. Putman and P. Düben, 'Global Cloud-Resolving Models', en, *Current Climate Change Reports* **5**, 172–184 (2019).
- ²³⁰J. Schlottke, W. Straub, K. D. Beheng, H. Gomaa and B. Weigand, 'Numerical Investigation of Collision-Induced Breakup of Raindrops. Part I: Methodology and Dependencies on Collision Energy and Eccentricity', en, *Journal of the Atmospheric Sciences* **67**, 557–575 (2010).
- ²³¹H. Schulz, R. Eastman and B. Stevens, 'Characterization and Evolution of Organized Shallow Convection in the Downstream North Atlantic Trades', en, *Journal of Geophysical Research: Atmospheres* **126**, e2021JD034575 (2021).
- ²³²H. Schulz and B. Stevens, 'Evaluating Large-Domain, Hecto-Meter, Large-Eddy Simulations of Trade-Wind Clouds Using EUREC⁴ A Data', en, *Journal of Advances in Modeling Earth Systems* **15**, e2023MS003648 (2023).
- ²³³J. Schwenkel, F. Hoffmann and S. Raasch, 'Improving collisional growth in Lagrangian cloud models: development and verification of a new splitting algorithm', en, *Geoscientific Model Development* **11**, 3929–3944 (2018).

- ²³⁴Segura, H. and Bayley, C. et al., 'A Single Tropical Rainbelt in Global Storm-Resolving Models: The Role of Surface Heat Fluxes Over the Warm Pool', en, *Journal of Advances in Modeling Earth Systems* **17**, 10.1029/2024MS004897 (2025).
- ²³⁵A. Seifert and T. Heus, 'Large-eddy simulation of organized precipitating trade wind cumulus clouds', en, *Atmospheric Chemistry and Physics* **13**, 5631–5645 (2013).
- ²³⁶A. Seifert, 'On the Parameterization of Evaporation of Raindrops as Simulated by a One-Dimensional Rainshaft Model', en, *Journal of the Atmospheric Sciences* **65**, 3608–3619 (2008).
- ²³⁷A. Seifert, A. Khain, U. Blahak and K. D. Beheng, 'Possible Effects of Collisional Breakup on Mixed-Phase Deep Convection Simulated by a Spectral (Bin) Cloud Model', en, *Journal of the Atmospheric Sciences* **62**, 1917–1931 (2005).
- ²³⁸A. Seifert, J. Leinonen, C. Siewert and S. Kneifel, 'The Geometry of Rimed Aggregate Snowflakes: A Modeling Study', en, *Journal of Advances in Modeling Earth Systems* **11**, 712–731 (2019).
- ²³⁹T. Seiki, C. Kodama, A. T. Noda and M. Satoh, 'Improvement in Global Cloud-System-Resolving Simulations by Using a Double-Moment Bulk Cloud Microphysics Scheme', en, *Journal of Climate* **28**, 2405–2419 (2015).
- ²⁴⁰T. A. Shaw and B. Stevens, 'The other climate crisis', en, *Nature* **639**, 877–887 (2025).
- ²⁴¹S. Shima, K. Kusano, A. Kawano, T. Sugiyama and S. Kawahara, '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: SUPER-DROPLET METHOD FOR CLOUDS AND PRECIPITATION', en, *Quarterly Journal of the Royal Meteorological Society* **135**, 1307–1320 (2009).
- ²⁴²S.-i. Shima, Y. Sato, A. Hashimoto and R. Misumi, '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', en, *Geoscientific Model Development* **13**, 4107–4157 (2020).
- ²⁴³B. J. Shipway and A. A. Hill, 'Diagnosis of systematic differences between multiple parametrizations of warm rain microphysics using a kinematic framework', en, *Quarterly Journal of the Royal Meteorological Society* **138**, 2196–2211 (2012).
- ²⁴⁴B. Shipway and A. Hill, 'The Kinematic Driver model (KiD)', Technical Re (2011).
- ²⁴⁵M. Simmel, T. Trautmann and G. Tetzlaff, 'Numerical solution of the stochastic collection equation—comparison of the Linear Discrete Method with other methods', en, *Atmospheric Research* **61**, 135–148 (2002).

- ²⁴⁶M. Simmel and S. Wurzler, 'Condensation and activation in sectional cloud microphysical models', *Atmospheric Research* **80**, 218–236 (2006).
- ²⁴⁷M. V. Smoluchowski, 'Drei Vorträge über Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen', *Zeitschrift für Physik* **17**, 557–585 (1916).
- ²⁴⁸A. M. Society (AMS), *Raindrop - Glossary of Meteorology*, English, Publication Title: raindrop - Glossary of Meteorology, Feb. 2025.
- ²⁴⁹I. Sölch and B. Kärcher, 'A large-eddy model for cirrus clouds with explicit aerosol and ice microphysics and Lagrangian ice particle tracking', en, *Quarterly Journal of the Royal Meteorological Society* **136**, 2074–2093 (2010).
- ²⁵⁰R. C. Srivastava, 'A Simple Model of Evaporatively Driven Downdraft: Application to Microburst Downdraft', en, *Journal of the Atmospheric Sciences* **42**, 1004–1023 (1985).
- ²⁵¹R. C. Srivastava, 'A Model of Intense Downdrafts Driven by the Melting and Evaporation of Precipitation', en, *Journal of the Atmospheric Sciences* **44**, 1752–1774 (1987).
- ²⁵²B. Stevens, 'Extreme Earth: Advancing global storm resolving models to usher in a new era of climate modelling and climate change science.', in, Vol. 20, Backup Publisher: EGU General Assembly 2018 Issue: EGU2018-10761 (2018).
- ²⁵³B. Stevens and S. Bony, 'What Are Climate Models Missing?', en, *Science* **340**, 1053–1054 (2013).
- ²⁵⁴B. Stevens, G. Feingold, W. R. Cotton and R. L. Walko, 'Elements of the Microphysical Structure of Numerically Simulated Nonprecipitating Stratocumulus', en, *Journal of the Atmospheric Sciences* **53**, 980–1006 (1996).
- ²⁵⁵B. Stevens, G. Vali, K. Comstock, R. Wood, M. C. Van Zanten, P. H. Austin, C. S. Bretherton and D. H. Lenschow, 'POCKETS OF OPEN CELLS AND DRIZZLE IN MARINE STRATOCUMULUS', en, *Bulletin of the American Meteorological Society* **86**, 51–58 (2005).
- ²⁵⁶B. Stevens, R. L. Walko, W. R. Cotton and G. Feingold, 'The Spurious Production of Cloud-Edge Supersaturations by Eulerian Models', en, *Monthly Weather Review* **124**, 1034–1041 (1996).
- ²⁵⁷B. Stevens et al., 'The Barbados cloud observatory: anchoring investigations of clouds and circulation on the edge of the itcz', *Bulletin of the American Meteorological Society* **97**, 787–801 (2016).
- ²⁵⁸B. Stevens et al., 'The Barbados Cloud Observatory: Anchoring Investigations of Clouds and Circulation on the Edge of the ITCZ', en, *Bulletin of the American Meteorological Society* **97**, 787–801 (2016).

- ²⁵⁹B. Stevens et al., 'A high-altitude long-range aircraft configured as a cloud observatory: the narval expeditions', *Bulletin of the American Meteorological Society* **100**, 1061–1077 (2019).
- ²⁶⁰B. Stevens et al., 'DYAMOND: the DYNamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains', en, *Progress in Earth and Planetary Science* **6**, 61 (2019).
- ²⁶¹B. Stevens et al., 'Sugar, gravel, fish and flowers: Mesoscale cloud patterns in the trade winds', en, *Quarterly Journal of the Royal Meteorological Society* **146**, 141–152 (2020).
- ²⁶²B. Stevens et al., 'The Added Value of Large-eddy and Storm-resolving Models for Simulating Clouds and Precipitation', en, *Journal of the Meteorological Society of Japan. Ser. II* **98**, 395–435 (2020).
- ²⁶³B. Stevens et al., 'EUREC⁴ A', en, *Earth System Science Data* **13**, 4067–4119 (2021).
- ²⁶⁴B. Stevens et al., 'ORCESTRA: Organized Convection and Earth-CARE Studies over the Tropical Atlantic', *Tellus submitted* (2026).
- ²⁶⁵B. Stevens and A. Seifert, 'Understanding macrophysical outcomes of microphysical choices in simulations of shallow cumulus convection', en, *Journal of the Meteorological Society of Japan. Ser. II* **86A**, 143–162 (2008).
- ²⁶⁶W. Straub, K. D. Beheng, A. Seifert, J. Schlottke and B. Weigand, 'Numerical Investigation of Collision-Induced Breakup of Raindrops. Part II: Parameterizations of Coalescence Efficiencies and Fragment Size Distributions', en, *Journal of the Atmospheric Sciences* **67**, 576–588 (2010).
- ²⁶⁷C.-Y. Su, W.-T. Chen, C.-M. Wu and H.-Y. Ma, 'Object-Based Evaluation of Tropical Precipitation Systems in DYAMOND Simulations over the Maritime Continent', en, *Journal of the Meteorological Society of Japan. Ser. II* **100**, 647–659 (2022).
- ²⁶⁸T. Suematsu, C. Kodama, Y. Yamada, H. Miura, D. Takasuka and T. Miyakawa, 'Microphysics dependency in 3.5km NICAM DYAMOND phase 2 experiments', in AGU Fall Meeting Abstracts, Vol. 2021, AGU Fall Meeting Abstracts (Dec. 2021), A45K–2001.
- ²⁶⁹L. Susskind and G. Hrabovsky, *The theoretical minimum: What you need to know to start doing physics* (Allen Lane, 2013).
- ²⁷⁰M. Szakáll and I. Urbich, 'Wind tunnel study on the size distribution of droplets after collision induced breakup of levitating water drops', en, *Atmospheric Research* **213**, 51–56 (2018).
- ²⁷¹D. Takasuka, M. Satoh, T. Miyakawa, C. Kodama, D. Klocke, B. Stevens, P. L. Vidale and C. R. Terai, 'A protocol and analysis of year-long simulations of global storm-resolving models and beyond', en, *Progress in Earth and Planetary Science* **11**, 66 (2024).

- ²⁷²D. Takasuka et al., 'How Can We Improve the Seamless Representation of Climatological Statistics and Weather Toward Reliable Global K-Scale Climate Simulations?', en, *Journal of Advances in Modeling Earth Systems* **16**, e2023MS003701 (2024).
- ²⁷³J. W. Telford, 'A NEW ASPECT OF COALESCENCE THEORY', en, *Journal of Meteorology* **12**, 436–444 (1955).
- ²⁷⁴F. Y. Testik, A. P. Barros and L. F. Bliven, 'Toward a Physical Characterization of Raindrop Collision Outcome Regimes', en, *Journal of the Atmospheric Sciences* **68**, 1097–1113 (2011).
- ²⁷⁵F. Y. Testik and M. K. Rahman, 'First in situ observations of binary raindrop collisions', en, *Geophysical Research Letters* **44**, 1175–1181 (2017).
- ²⁷⁶F. Y. Testik, 'Outcome regimes of binary raindrop collisions', en, *Atmospheric Research* **94**, 389–399 (2009).
- ²⁷⁷J. Tian, Y. Zhang, S. A. Klein, C. R. Terai, P. M. Caldwell, H. Beydoun, P. Bogenschutz, H.-Y. Ma and A. S. Donahue, 'How Well Does the DOE Global Storm Resolving Model Simulate Clouds and Precipitation Over the Amazon?', en, *Geophysical Research Letters* **51**, e2023GL108113 (2024).
- ²⁷⁸J. R. R. Tolkien, *The Return of the King* (George Allen and Unwin, London, 1955).
- ²⁷⁹H. Tomita, H. Miura, S. Iga, T. Nasuno and M. Satoh, 'A global cloud-resolving simulation: Preliminary results from an aqua planet experiment', en, *Geophysical Research Letters* **32**, 2005GL022459 (2005).
- ²⁸⁰H. Tomita and M. Satoh, 'A new dynamical framework of nonhydrostatic global model using the icosahedral grid', *Fluid Dynamics Research* **34**, 357–400 (2004).
- ²⁸¹L. Touzé-Peiffer, R. Vogel and N. Rochetin, 'Cold Pools Observed during EUREC4A: Detection and Characterization from Atmospheric Soundings', EN, *Journal of Applied Meteorology and Climatology* **61**, 593–610 (2022).
- ²⁸²F. Tridon, A. Battaglia and D. Watters, 'Evaporation in action sensed by multiwavelength Doppler radars', en, *Journal of Geophysical Research: Atmospheres* **122**, 9379–9390 (2017).
- ²⁸³C. Trott, L. Berger-Vergiat, D. Poliakoff, S. Rajamanickam, D. Lebrun-Grandie, J. Madsen, N. Al Awar, M. Gligoric, G. Shipman and G. Womeldorff, 'The Kokkos EcoSystem: Comprehensive Performance Portability for High Performance Computing', *Computing in Science Engineering* **23**, 10–18 (2021).
- ²⁸⁴C. R. Trott et al., 'Kokkos 3: Programming Model Extensions for the Exascale Era', *IEEE Transactions on Parallel and Distributed Systems* **33**, 805–817 (2022).

- ²⁸⁵S. Unterstrasser and I. Sölch, ‘Optimisation of the simulation particle number in a Lagrangian ice microphysical model’, en, *Geoscientific Model Development* **7**, 695–709 (2014).
- ²⁸⁶S. Unterstrasser, F. Hoffmann and M. Lerch, ‘Collection/aggregation algorithms in Lagrangian cloud microphysical models: rigorous evaluation in box model simulations’, en, *Geoscientific Model Development* **10**, 1521–1548 (2017).
- ²⁸⁷S. Unterstrasser, F. Hoffmann and M. Lerch, ‘Collisional growth in a particle-based cloud microphysical model: insights from column model simulations using LCM1D (v1.0)’, en, *Geoscientific Model Development* **13**, 5119–5145 (2020).
- ²⁸⁸V.C.A.F., ‘The Mathematical Theory of Non-uniform Gases. An account of the kinetic theory of viscosity, thermal conduction, and diffusion in gases. By Sydney Chapman and T. G. Cowling. 2nd Edn., pp. ix, 431. 60s. 1952. (Cambridge University Press)’, *The Mathematical Gazette* **38**, 63–64 (1954).
- ²⁸⁹M. C. vanZanten et al., ‘Controls on precipitation and cloudiness in simulations of trade-wind cumulus as observed during RICO: CONTROLS ON PRECIPITATION ... IN TRADE-WIND CUMULUS’, en, *Journal of Advances in Modeling Earth Systems* **3**, n/a–n/a (2011).
- ²⁹⁰A. Varble, A. M. Fridlind, E. J. Zipser, A. S. Ackerman, J.-P. Chaboureaud, J. Fan, A. Hill, S. A. McFarlane, J.-P. Pinty and B. Shipway, ‘Evaluation of cloud-resolving model intercomparison simulations using TWP-ICE observations: Precipitation and cloud structure’, en, *Journal of Geophysical Research* **116**, D12206 (2011).
- ²⁹¹A. Varble, E. J. Zipser, A. M. Fridlind, P. Zhu, A. S. Ackerman, J.-P. Chaboureaud, S. Collis, J. Fan, A. Hill and B. Shipway, ‘Evaluation of cloud-resolving and limited area model intercomparison simulations using TWP-ICE observations: 1. Deep convective updraft properties’, en, *Journal of Geophysical Research: Atmospheres* **119**, 10.1002/2013JD021371 (2014).
- ²⁹²A. Varble, E. J. Zipser, A. M. Fridlind, P. Zhu, A. S. Ackerman, J.-P. Chaboureaud, J. Fan, A. Hill, B. Shipway and C. Williams, ‘Evaluation of cloud-resolving and limited area model intercomparison simulations using TWP-ICE observations: 2. Precipitation microphysics’, en, *Journal of Geophysical Research: Atmospheres* **119**, 10.1002/2013JD021372 (2014).
- ²⁹³J. Vial, A. L. Albright, R. Vogel, I. Musat and S. Bony, ‘Cloud transition across the daily cycle illuminates model responses of trade cumuli to warming’, en, *Proceedings of the National Academy of Sciences* **120**, e2209805120 (2023).

- ²⁹⁴J. Vial, S. Bony, B. Stevens and R. Vogel, 'Mechanisms and Model Diversity of Trade-Wind Shallow Cumulus Cloud Feedbacks: A Review', en, *Surveys in Geophysics* **38**, 1331–1353 (2017).
- ²⁹⁵J. Vial, R. Vogel, S. Bony, B. Stevens, D. M. Winker, X. Cai, C. Hohenegger, A. K. Naumann and H. Brogniez, 'A New Look at the Daily Cycle of Trade Wind Cumuli', en, *Journal of Advances in Modeling Earth Systems* **11**, 3148–3166 (2019).
- ²⁹⁶J. Vial, R. Vogel and H. Schulz, 'On the daily cycle of mesoscale cloud organization in the winter trades', en, *Quarterly Journal of the Royal Meteorological Society* **147**, 2850–2873 (2021).
- ²⁹⁷R. Vogel, A. L. Albright, J. Vial, G. George, B. Stevens and S. Bony, 'Strong cloud–circulation coupling explains weak trade cumulus feedback', en, *Nature*, [10.1038/s41586-022-05364-y](https://doi.org/10.1038/s41586-022-05364-y) (2022).
- ²⁹⁸R. Vogel, H. Konow, H. Schulz and P. Zuidema, 'A climatology of trade-wind cumulus cold pools and their link to mesoscale cloud organization', en, *Atmospheric Chemistry and Physics* **21**, 16609–16630 (2021).
- ²⁹⁹R. Vogel, L. Nuijens and B. Stevens, 'The role of precipitation and spatial organization in the response of trade-wind clouds to warming', en, *Journal of Advances in Modeling Earth Systems* **8**, 843–862 (2016).
- ³⁰⁰U. Wacker and A. Seifert, 'Evolution of rain water profiles resulting from pure sedimentation: Spectral vs. parameterized description', en, *Atmospheric Research* **58**, 19–39 (2001).
- ³⁰¹N. Wedi, P. Bauer, W. Deconinck, M. Diamantakis, M. Hamrud, C. Kuehnlein, S. Malardel, K. Mogensen, G. Mozdzyński and P. Smolarkiewicz, *The modelling infrastructure of the Integrated Forecasting System: Recent advances and future challenges*, eng, Issue: 760 Publication Title: ECMWF Technical Memoranda, 2015.
- ³⁰²T. Wehr et al., 'The earthcare mission – science and system overview', *Atmospheric Measurement Techniques* **16**, 3581–3608 (2023).
- ³⁰³J.-N. Welss, C. Siewert and A. Seifert, 'Explicit habit-prediction in the Lagrangian super-particle ice microphysics model McSnow', Publisher: Wiley, [10.22541/essoar.168614461.18006193/v1](https://doi.org/10.22541/essoar.168614461.18006193/v1) (2023).
- ³⁰⁴E. M. Wilcox and L. J. Donner, 'The Frequency of Extreme Rain Events in Satellite Rain-Rate Estimates and an Atmospheric General Circulation Model', en, *Journal of Climate* **20**, 53–69 (2007).
- ³⁰⁵J. D. Wille, R. Koch, T. Becker and E. Fischer, 'Extreme Precipitation Depiction in Convection-Permitting Earth System Models Within the nextGEMS Project', en, *Journal of Advances in Modeling Earth Systems* **17**, e2024MS004840 (2025).

- ³⁰⁶M. Wirth, A. Fix, P. Mahnke, H. Schwarzer, F. Schrandt and G. Ehret, 'The airborne multi-wavelength water vapor differential absorption lidar wales: system design and performance', *Applied Physics B* **96**, 201–213 (2009).
- ³⁰⁷L. Xue et al., 'Idealized Simulations of a Squall Line from the MC3E Field Campaign Applying Three Bin Microphysics Schemes: Dynamic and Thermodynamic Structure', en, *Monthly Weather Review* **145**, 4789–4812 (2017).
- ³⁰⁸C. Yin, S.-i. Shima, L. Xue and C. Lu, 'Simulation of marine stratocumulus using the super-droplet method: numerical convergence and comparison to a double-moment bulk scheme using SCALE-SDM 5.2.6-2.3.1', en, *Geoscientific Model Development* **17**, 5167–5189 (2024).
- ³⁰⁹G. Zängl, D. Reinert, P. Rípodas and M. Baldauf, 'The *ICON* (ICOsahedral Non-hydrostatic) modelling framework of *DWD* and *MPI-M* : Description of the non-hydrostatic dynamical core', en, *Quarterly Journal of the Royal Meteorological Society* **141**, 563–579 (2015).
- ³¹⁰R. Zhang, L. Zhou, S.-i. Shima and H. Yang, 'Preliminary evaluation of the effect of electro-coalescence with conducting sphere approximation on the formation of warm cumulus clouds using SCALE-SDM version 0.2.5–2.3.0', en, *Geoscientific Model Development* **17**, 6761–6774 (2024).
- ³¹¹P. Zmijewski, P. Dziekan and H. Pawlowska, 'Modeling Collision-Coalescence in Particle Microphysics: Numerical Convergence of Mean and Variance of Precipitation in Cloud Simulations Using University of Warsaw Lagrangian Cloud Model (UWLCM) 2.', en, (2023).
- ³¹²P. Zuidema, Z. Li, R. J. Hill, L. Bariteau, B. Rilling, C. Fairall, W. A. Brewer, B. Albrecht and J. Hare, 'On Trade Wind Cumulus Cold Pools', en, *Journal of the Atmospheric Sciences* **69**, 258–280 (2012).
- ³¹³P. Zuidema, G. Torri, C. Muller and A. Chandra, 'A Survey of Precipitation-Induced Atmospheric Cold Pools over Oceans and Their Interactions with the Larger-Scale Environment', en, in *Shallow Clouds, Water Vapor, Circulation, and Climate Sensitivity*, edited by R. Pincus, D. Winker, S. Bony and B. Stevens (Springer International Publishing, Cham, 2017), pp. 111–133.



Photos taken during the [ORCESTRA](#) campaign by anonymous or this author, Cabo Verde and Barbados, August – September 2024.

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst’s seminal book on typography “*The Elements of Typographic Style*”. `classicthesis` is available for both $\text{\LaTeX}$ and $\text{\LyX}$:

<https://bitbucket.org/amiede/classicthesis/>

Final Version as of 19th January 2026.

DECLARATION

Eidesstattliche Versicherung

Declaration upon Oath

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Sofern im Zuge der Erstellung der vorliegenden Dissertationsschrift generative Künstliche Intelligenz (gKI) basierte elektronische Hilfsmittel verwendet wurden, versichere ich, dass meine eigene Leistung im Vordergrund stand und dass eine vollständige Dokumentation aller verwendeten Hilfsmittel gemäß der Guten wissenschaftlichen Praxis vorliegt. Ich trage die Verantwortung für eventuell durch die gKI generierte fehlerhafte oder verzerrte Inhalte, fehlerhafte Referenzen, Verstöße gegen das Datenschutz- und Urheberrecht oder Plagiate.

I hereby declare and affirm that this doctoral dissertation is my own work and that I have not used any aids and sources other than those indicated.

If electronic resources based on generative artificial intelligence (gAI) were used in the course of writing this dissertation, I confirm that my own work was the main and value-adding contribution and that complete documentation of all resources used is available in accordance with good scientific practice. I am responsible for any erroneous or distorted content, incorrect references, violations of data protection and copyright law or plagiarism that may have been generated by the gAI.

Hamburg, 16th January 2026

Clara J.A. Bayley

Eidesstattliche Versicherung

Declaration upon Oath

Ich versichere, dass dieses gebundene Exemplar der Dissertation und das in elektronischer Form eingereichte Dissertationsexemplar (über den Docata-Upload) und das bei der Fakultät zur Archivierung eingereichte gedruckte gebundene Exemplar der Dissertationsschrift identisch sind.

I, the undersigned, declare that this bound copy of the dissertation and the dissertation submitted in electronic form (via the Docata upload) and the printed bound copy of the dissertation submitted to the faculty (responsible Academic Office or the Doctoral Office Physics) for archiving are identical.

Hamburg, 16th January 2026

Clara J.A. Bayley

Hinweis / Reference

Die gesamten Veröffentlichungen in der Publikationsreihe des MPI-M
„Berichte zur Erdsystemforschung / Reports on Earth System Science“,
ISSN 1614-1199

sind über die Internetseiten des Max-Planck-Instituts für Meteorologie erhältlich:
<https://mpimet.mpg.de/forschung/publikationen>

*All the publications in the series of the MPI -M
„Berichte zur Erdsystemforschung / Reports on Earth System Science“,
ISSN 1614-1199*

*are available on the website of the Max Planck Institute for Meteorology:
<https://mpimet.mpg.de/en/research/publications>*

