https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

A Lagrangian Particle Tracking Framework for the Super-Droplet Method: Development, Implementation, and Application of Backward and Forward Algorithms in SCALE-SDM 5.2.6-2.3.1

Chongzhi Yin[1], Shin-ichiro Shima[2, 3], Chunsong Lu[1]

[1]China Meteorological Administration Aerosol-Cloud and Precipitation Key Laboratory/Collaborative Innovation Centre on Forecast and Evaluation of Meteorological Disasters (CIC-FEMD), Nanjing University of Information Science and Technology, Nanjing 210044, China

[2]Graduate School of Information Science, University of Hyogo, Kobe 6500047, Japan

[3]RIKEN Center for Computational Science, Kobe 6500047, Japan

Correspondence to: Shin-ichiro Shima (s_shima@gsis.u-hyogo.ac.jp) and Chunsong Lu (luchunsong110@163.com)

Abstract. Understanding the complete lifecycle of cloud hydrometeors is fundamental to advancing cloud microphysics, yet a formally documented and computationally efficient framework for tracking individual particles through complex processes like coalescence within parallelized cloud models has been lacking. This study addresses this methodological gap by presenting the detailed design, implementation, and application of two complementary super-droplet tracking algorithms—a backward "lineage" tracking algorithm and a forward "tagging" tracking algorithm—developed within a Large-Eddy Simulation model coupled with the Super-Droplet Method (SDM). The backward algorithm establishes a direct historical link for every super-droplet, enabling efficient $O(1)$ lookup for reconstructing a particle's complete microphysical history. The forward algorithm employs a stratified random sampling method to select and assign persistent identifiers to a representative cohort of super-droplets, allowing for detailed prognostic analysis with manageable data storage costs. A key feature of both algorithms is the comprehensive method for recording and outputting detailed information on coalescence events. The utility and power of these algorithms are demonstrated in a case study of marine stratocumulus. The framework enabled a quantitative, process-level validation of the critical 15-20 μm radius range for the onsite of efficient warm rain initiation. Furthermore, the lineage-tracing capability mechanistically confirmed the classical formation pathway of large droplets within cloud-base updrafts, directly linking large-scale turbulent structures to the lifecycle of individual precipitation embryos. In conclusion, the tracking algorithms presented here provide the scientific community with a powerful and versatile toolset to investigate the intricate lifecycle of cloud particles with unprecedented detail and offer a robust methodology for evaluating and improving microphysical parameterizations in larger-scale weather and climate models.

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

1 Introduction

Tracking the evolutionary history of individual objects is a fundamental technique in scientific research, with mature methodologies developed across various fields such as astrophysics, meteorology, fluid dynamics, and computer science. In atmospheric science, Lagrangian particle dispersion models are a key tool for elucidating the complex pathways of atmospheric constituents. For instance, well-established trajectory models like HYSPLIT (Draxler and Hess, 1998; Stein et al., 2015) and FLEXPART (Stohl et al., 2005; Pisso et al., 2019) are widely used for critical applications ranging from predicting the dispersion of volcanic ash and industrial pollutants to quantifying the moisture sources for major precipitation events. These models effectively answer the questions of "where will it go?" (forward tracking) and "where did it come from?" (backward tracking) for air parcels on a macroscopic scale.

In parallel, computer science has formalized this concept into the general fields of "data provenance" or "lineage tracking", which focus on recording all transformations an object undergoes from its creation to its current state (Simmhan et al., 2005; Freire et al., 2008). In essence, data provenance acts as a detailed digital laboratory notebook for a complex simulation. For any given result, it provides a verifiable record of the inputs, processes, and intermediate steps that produced it, making it a critical technology for ensuring reproducibility, debugging, and interpretability in large-scale simulations and data-intensive computing.

While these tracking and provenance concepts are well-established at macroscopic and theoretical levels, their application to the intricate world of cloud microphysics presents unique challenges and opportunities. The Super-Droplet Method (SDM), a powerful particle-based Lagrangian approach, has emerged as a state-of-the-art technique for simulating clouds with detailed microphysics, where each super-droplet represents a multitude of real particles (Shima et al., 2009; Shima et al., 2020). Although recent studies using the SDM have presented the trajectories of cloud droplets and ice crystals as compelling scientific outcomes (e.g., Chandrakar et al. (2022); Chandrakar et al. (2024)), the detailed design, implementation, and performance of the underlying tracking algorithms themselves have not been formally published.

Alternative approaches for studying cloud processes also exist, but each comes with distinct limitations. Parcel models, for instance, track the motion and thermodynamic properties of a bulk air parcel and often rely on simplified assumptions such as adiabatic expansion and internal homogeneity (e.g., Squires (1958); Krueger et al. (1997)), thereby neglecting the individual histories and turbulent interactions of the droplets within. In contrast, the SDM framework allows for a more detailed examination of processes like condensation and collision-coalescence for individual super-droplets under realistic, turbulent conditions. Bin microphysics schemes track the evolution of the droplet size distribution by discretizing it into size bins (e.g., Khain et al. (2000); Khain et al. (2015)). However, this approach is limited by the resolution of the bins, where important microphysical details may be averaged out, and numerical diffusion can artificially broaden the spectrum. The SDM avoids these limitations by continuously following individual super-droplets, allowing for a more precise representation of their microphysical interactions without relying on binning. At the highest fidelity, Direct Numerical Simulation (DNS) is capable of tracking individual particles with exceptional resolution (e.g., Wang and Maxey (1993);

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

Ayala et al. (2008a); Ayala et al. (2008b)), but its immense computational cost restricts its application to small domain sizes and low Reynolds numbers. This makes DNS impractical for studying cloud-scale structures where large eddies are crucial. The coupling of the SDM with a Large-Eddy Simulation (LES) framework overcomes this by enabling the study of cloud microphysics in larger, more realistic domains, capturing both small-scale interactions and large-scale turbulent dynamics with computational efficiency. Therefore, a general-purpose, computationally efficient, and formally documented tracking algorithm capable of handling complex microphysical processes within a massively parallel SDM framework is still lacking in the literature.

Recent advancements have pushed the boundaries of SDM simulations to unprecedented scales. For example, the work by Matsushima et al. (2023) demonstrated the feasibility of meter- to submeter-scale simulations by focusing on profound computational optimizations for high-performance computing systems. To validate their improved velocity interpolation scheme, their study employed a simple forward tracking approach, assigning a unique identifier to each super-droplet at the initial time. As they described, "By using the ID for analysis, we calculated the initial position of the SD to investigate SD mixing", which was then visualized by colouring particles based on their initial coordinates (e.g., their Figure 6). While this effectively underscores the scientific value of particle-based analysis for visualizing complex processes like chaotic mixing, the primary objective of their paper was to overcome computational performance barriers, the formal design and detailed implementation of a general-purpose particle tracking methodology—particularly one capable of full lineage reconstruction through coalescence via backward tracking, or one designed for detailed process analysis beyond visualization—was not a central component of their contribution. Therefore, a robust, computationally efficient, and formally documented tracking algorithm designed for broad scientific inquiry within massively parallel SDM frameworks is still a critical missing element in the literature.

This paper aims to fill this methodological gap by presenting the development, implementation, and application of two distinct super-droplet tracking algorithms—a backward "lineage" tracking and a forward "tagging" tracking—within an LES framework coupled with the SDM. The importance of this contribution is significant, as understanding the individual histories of hydrometeors is crucial for fundamental questions in cloud physics, from condensational growth to precipitation initiation. Our backward tracking algorithm provides an efficient method to analyse the detailed microphysical histories of specific targets, such as large cloud droplets, while the forward tracking algorithm enables statistically robust analysis of a representative cohort of particles with manageable data storage costs. By detailing these algorithms, which include a comprehensive method for recording and outputting detailed information on coalescence events, we provide the atmospheric science community with a versatile and powerful toolset. Other researchers can apply these methods to analyse the origins and fates of cloud particles in various cloud regimes or use the insights gained to inform and improve microphysical parameterizations in larger-scale models.

The remainder of this paper is organized as follows. Section 2 provides a detailed description of the methodology and implementation of the two complementary tracking algorithms—the backward "lineage" tracking and the forward "tagging" tracking—within the SCALE-SDM framework, including the post-processing workflows for each. In Section 3, we

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

demonstrate the utility and power of these algorithms through their application to a classic marine stratocumulus case study (DYCOMS-II RF02), presenting key scientific insights gained from both the forward and backward analyses. Section 4 discusses these findings in a broader context, synthesizes the strengths and limitations of the developed framework, and outlines avenues for future work. Finally, Section 5 summarizes our main conclusions and reiterates the significance of this contribution to the cloud physics community.

2 The Tracking Algorithms: Methodology and Implementation

2.1 General Framework: Numerical Model

The forward and backward tracking algorithms presented in this study were developed and implemented within the framework of the Scalable Computing for Advanced Library and Environment (SCALE) model, version 5.2.6 (Sato et al., 2015; Nishizawa et al., 2015). SCALE is an open-source software library jointly developed by institutions including the RIKEN (Institute of Physical and Chemical Research) Center for Computational Science and the Japan Meteorological Agency. It is designed to provide a flexible, efficient, and scalable numerical simulation platform for weather and climate research. Key features of the SCALE framework include a modular design, allowing users to combine various physical schemes to suit different research needs, and robust support for parallel computing (including MPI, OpenMP and OpenACC) to ensure high performance on diverse computational architectures.

For this work, we utilize SCALE-RM, the regional model component of the framework, whose dynamical core is based on the non-hydrostatic, compressible fluid equations. The model is configured as an LES, which explicitly resolves large-scale turbulent motions while parameterizing unresolved small-scale turbulence using a sub-grid scale (SGS) model. This dynamical core is coupled with the SDM, forming the SCALE-SDM version 5.2.6-2.3.1 (Shima et al., 2009; Yin et al., 2024), configuration used in this study. This coupling allows for the detailed investigation of cloud-scale dynamics and their interaction with high-resolution microphysical processes.

A critical feature of the model's architecture, which directly impacts the design of any particle-based algorithm, is its parallelization strategy. To enable large-scale simulations on high-performance computing systems, SCALE-SDM employs a spatial domain decomposition method using the Message Passing Interface (MPI). As illustrated by the model documentation, the total simulation domain is partitioned into multiple sub-domains, with each sub-domain assigned to a separate MPI process for computation.

This distributed-memory architecture imposes a fundamental requirement on our tracking algorithms: they must robustly handle the transfer of super-droplets as they move across the boundaries of these MPI sub-domains. Consequently, a significant component of our algorithm design, detailed in subsequent sections, is dedicated to ensuring the continuity and integrity of particle tracking information (both historical and permanent identifiers) during these inter-process communication events.

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

2.2 The Backward Tracking

2.2.1 The Backward Tracking Algorithm Description

The objective of the backward tracking algorithm is to enable the detailed, time-resolved reconstruction of a super-droplet's complete microphysical history and trajectory from a known final state. This capability is essential for performing source attribution and process analysis, for instance, to investigate the formation mechanisms of large raindrops or to understand the environmental history of particles that end up in specific locations. The methodology is formally presented in its high-level structure in

Algorithm 1. The full implementation of the algorithm is publicly available in the archived source code (Yin, 2025b). A central tenet of its design is computational efficiency; the algorithm establishes a direct historical link for each super-droplet, which facilitates highly efficient lookups during post-processing, rather than relying on a computationally expensive search-based method.

The process begins with the INITIALIZE_SD_BACKWARD_TRACKING procedure, which is invoked once at the start of the simulation. This procedure initializes the historical identifier arrays, pre_sdid and pre_dmid, to a predefined INVALID_ID value. These two variables are fundamental to the tracking logic. For any given super-droplet, pre_dmid stores the identifier of the MPI sub-domain where the particle was located at the previous output time, while pre_sdid stores the sequence number of that super-droplet within its corresponding output file for that sub-domain. Initializing these arrays ensures that all super-droplets at the initial time step (t = 0) have no antecedent history.

The main logic is encapsulated within the ADVANCE_ONE_TIME_STEP_BACKWARD procedure, which executes at every model time step. The sequence of operations within this procedure is critical for maintaining the integrity of the historical tracking chain, following a "use-first, then-update" paradigm, and can be broken down into the following phases:

(1) Phase 1: Calculation of Physics. The physical evolution of all super-droplets is calculated first. This phase includes the SD_MOTION_TRACKING subroutine to update super-droplet positions based on the wind fields and sedimentations, the SD_BOUNDARY_HANDLING subroutine to manage particles crossing MPI domain boundaries and ensure tracking continuity, and crucially, the SD_COALESCENCE_TRACKING subroutine. This subroutine reads from the pre_sdid and pre_dmid arrays, which hold the historical information from the end of the previous time step, to identify the parent super-droplets involved in a collision. It then sets the if_coal flag to 1. This flag serves as an indicator for the post-processing algorithm to retrieve the detailed collision attributes (e.g., the IDs of the colliding pair) from the high-resolution coalescence event log, which records every event sequentially at the microphysical time step (dt_mp). It should be noted that other standard microphysical processes, such as activation/condensation and deactivation/evaporation which modify droplet properties but do not alter the historical identifiers, are also computed within this phase but are omitted from the high-level description for clarity.

(2) Phase 2: Data Output. After the physical state for the current time step has been fully determined, the OUTPUT_SD_STATE_DATA procedure is called at user-defined output intervals. This procedure writes the historical

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

identifiers (pre_sdid, pre_dmid) and the current coalescence status (if_coal) to the output file, providing a comprehensive record of the system state and super-droplet lineage at key temporal snapshots.

(3) Phase 3: History Update for Next Step. The final and most critical step for the backward tracking logic is the UPDATE_HISTORICAL_IDS procedure, executed at the end of every time step. This procedure assigns the definitive identifiers of each super-droplet at the end of the current time step (current_sdid, current_dmid) to the historical tracking arrays (pre_sdid, pre_dmid). It also resets the if_coal flag to zero. This action effectively prepares the historical state needed for the subsequent time step's calculations, ensuring the continuity of the lineage chain.

A more detailed look into the SD_COALESCENCE_TRACKING procedure reveals the core of the lineage recording. Within this subroutine, when a collision between two super-droplets is determined to occur, the algorithm first records the state of the two parent super-droplets before their properties (e.g., radius and multiplicity) are updated by the coalescence. Specifically, it reads the pre_sdid and pre_dmid from the main historical arrays to identify the parent super-droplets, and also records their current radius and multiplicity (where multiplicity refers to the number of real droplets represented by a single super-droplet). This information is immediately written to a dedicated coalescence event file at each microphysics time step. To clarify the data flow across different time scales, a schematic illustration of the lineage recording mechanism is provided in Figure 1. As shown in the timeline, the model integrates microphysics at a high-frequency time step (dt_mp), while the full system state is output at a lower-frequency interval (dt_out). Even if a super-droplet experiences multiple coalescence events within a single output interval (e.g., at t_i and t_{i+1}), each event is instantly recorded in a dedicated event log file. The if_coal flag stored in the main state file at t_{out} serves as a diagnostic index, allowing the backward tracking algorithm to accurately retrieve these intermediate events from the logs and reconstruct the unbroken lineage chain during post-processing.

The model also calculates the total number of coalescence events between the pair. Referring the stochastic collection algorithm described in Shima et al. (2009), when the collision probability (p_α) exceeds one, it is interpreted as the expected number of collection events within the time step. A random integer (γ_α) is then generated, whose mathematical expectation equals p_α, to represent the actual number of collisions performed. Following the physical coalescence, the properties of the participating super-droplets are updated according to the model's rules. A special condition arises when both colliding super-droplets have a multiplicity of one. In this scenario, the smaller of the two is consumed and removed from the simulation, and its tracking identifiers (pre_sdid, pre_dmid) can be considered invalid for any subsequent steps, thus terminating its historical lineage. In all other cases, both parent super-droplets are preserved with updated properties. By capturing this comprehensive pre-collision snapshot, the algorithm ensures that a complete and accurate historical record is maintained for every coalescence event. The computational advantage of this design is realized during post-processing, where tracing a super-droplet's history back through time becomes a series of $O(1)$ direct lookups, rather than an $O(N)$ search at each step.

It is important to note, however, that the main state output, which includes these historical linkage variables, occurs at user-defined intervals (output_time) that may be longer than the model's integration time step and microphysics time step. In our

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

implementation, to balance the need for historical tracking with practical data storage and I/O constraints, the pre_sdid and pre_dmid arrays are continuously updated at every time step, but the full state, including this linkage, is only written to file during these output steps. This design makes long-term simulations feasible, though it implies that post-processing analysis of the particle lineage between output frames relies on the detailed coalescence event logs.

2.2.2 Post-processing for Backward Tracking

The data generated by the backward tracking algorithm enables a detailed, time-resolved reconstruction of a super-droplet's lineage from a known final state. The post-processing workflow is designed to efficiently trace these histories from a large and distributed dataset, which is typically split by both MPI sub-domain and output time.

The process begins by identifying a set of "seed" super-droplets at a specified end time of the analysis period. These seeds are selected based on user-defined physical criteria, such as exceeding a certain radius or being located in a region of interest.

Once the initial seed particles are identified, the core of the post-processing script begins tracing their histories backward in time.

To balance memory usage and I/O bottlenecks, the tracing is performed in batches. Instead of attempting to trace all seed particles simultaneously, a small batch of trajectories is processed at once. For each active trajectory within a batch, the script uses the stored historical identifiers, pre_dmid and pre_dmid, to determine the exact data file and the specific record index of the parent super-droplet at the preceding output time step. This direct-lookup mechanism, which has an $\mathcal{O}(1)$ complexity, is fundamental to the efficiency of the post-processing, as it entirely avoids computationally expensive searches. To further optimize I/O, required data files for the current time step are loaded into a temporary in-memory cache, preventing redundant disk reads if multiple trajectories within the batch trace back to the same file.

The script iteratively repeats this lookup process, moving backward from one time step to the next and recording the state of the parent particle at each step. If a traced super-droplet has its if_coal flag set to 1, this indicates it was formed by a coalescence event. The post-processing workflow can then reference the separate coalescence event logs to identify the two "grandparent" particles, allowing the lineage to branch and be traced further. This iterative process continues until the beginning of the simulation is reached or the lineage terminates (e.g., the pre_sdid becomes invalid). Finally, the reconstructed trajectories from all batches are collated and written to a single, analysis-ready NetCDF file, where each particle's complete history is stored in a consistent format.

2.3 The Forward Tracking

2.3.1 The Forward Tracking Algorithm Description

The forward tracking algorithm is designed to follow the evolution of a pre-selected, representative subset of super-droplets from the initial time of the simulation. Its objective is to reconstruct the future trajectories and microphysical histories of these specific particles, allowing for a detailed statistical analysis of a cohort's behaviour over time. The methodology is

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

formally presented in its high-level structure in Algorithm 2. The full implementation of the algorithm is publicly available in the archived source code (Yin, 2025a). This is achieved by assigning a persistent, immutable identifier to each selected super-droplet, which serves as its unique tag throughout its lifecycle.

The process commences with the INITIALIZE_SD_FORWARD_TRACKING procedure, which follows a distinct two-stage logic. First, the permanent tracking identifiers, sd_id and dm_id, are initialized to an INVALID_ID value for all super-droplets in the simulation, ensuring that no particles are tracked by default. For a tracked super-droplet, sd_id represents its permanent unique identifier, while dm_id stores the identifier of the MPI sub-domain where the super-droplet was initially selected; both identifiers remain constant unless the particle is lost. The second stage, encapsulated in the SELECT_AND_TAG_PARTICLES subroutine, then executes. This subroutine employs a stratified random sampling strategy to select a representative subset of super-droplets based on physical criteria, such as user-defined height layers and radius thresholds, and assigns the unique tracking identifiers to them.

Once initialized, the core of the algorithm resides in the ADVANCE_ONE_TIME_STEP_FORWARD procedure, which executes at every model time step. The following phases are included:

(1) Phase 1: Calculation of Physics. The physical evolution for the entire ensemble of super-droplets is calculated first. This includes updating their spatial coordinates via the SD_MOTION_TRACKING subroutine, which accounts for advection by the resolved wind fields. Additionally, the ADD_RANDOM_PERTURBATION_ALL procedure is called to apply small stochastic perturbations to the motion of each super-droplet, representing sub-grid scale turbulent effects. Following this, the SD_BOUNDARY_HANDLING procedure manages super-droplets crossing MPI domain boundaries, ensuring that the persistent sd_id and dm_id of any tagged super-droplets are passed along with their physical properties to maintain tracking continuity. The SD_COALESCENCE_TRACKING subroutine then identifies all collision events. Critically, it checks if at least one of the colliding super-droplets is a tracked particle (i.e., its sd_id is not INVALID_ID). Only if a tagged super-droplet is involved is the event's information recorded for output. This procedure also handles the termination of tracking for a tagged super-droplet if it is consumed during a coalescence event, marking it as “lost” by resetting its sd_id to INVALID_ID.

(2) Phase 2: Data Output and Flag Reset. At user-defined output intervals, the OUTPUT_TRACKED_SD_DATA procedure is called. This procedure selectively writes the full state variables of only the tagged and still-valid super-droplets to the output files. The identity of these super-droplets is confirmed by checking that their sd_id is not equal to INVALID_ID. Following the data output, the RESET_COALESCENCE_FLAGS procedure is called to reset the if_coal flag for all super-droplets to zero in preparation for the next output interval.

2.3.2 Post-processing for Forward Tracking

The output from the forward tracking algorithm consists of time-series data for a specific cohort of tagged super-droplets, typically distributed across numerous files split by both MPI sub-domain and output time. The primary goal of post-

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

processing is to reconstruct the complete lifecycle of each individually tagged super-droplet from this distributed dataset into a single, analysis-ready file. The workflow involves several key steps.

First, for each output time step, the data from all MPI-split files are read and concatenated into a single set of arrays. To establish a consistent record for each particle across all time steps, this aggregated data is then sorted based on a composite key of the unique, permanent (dm_id, sd_id) pair. This ensures that the super-droplet at a specific index k in the array at one time step is the same super-droplet as the one at index k at any other time step.

A critical challenge in forward tracking is handling particles that are "lost" during the simulation, for instance, by being consumed in a coalescence event. The post-processing workflow addresses this by first establishing a master list of all tracked particles from the initial time step. Then, for each subsequent time step, it identifies which particles from this master list are missing. To maintain a consistent data structure, a full-sized array matching the initial particle count is created. Data for the surviving particles are placed into their original indexed positions, while the array slots corresponding to the lost particles are filled with a defined invalid value (e.g., NaN or -9999).

This process is repeated for all time steps, and the resulting consistently-ordered and full-sized arrays are collated into a final data structure, such as a single NetCDF file. This approach, which loads the data for all tracked super-droplets into memory at once, is effective and manageable when the number of selected particles is moderate, as is the case in this study. However, for scenarios involving a significantly larger number of tracked super-droplets, this method could become memory-intensive. In such cases, alternative strategies would be necessary, such as processing the data in a streaming, time-step-by-time-step fashion to reduce the memory footprint, or utilizing a database to store and query particle histories more efficiently. Finally, when interpreting domain-wide statistics derived from the tracked cohort, it is important to consider that the spatial distribution of the particles may become uneven over long simulation times, which could affect their representativeness of the entire cloud field.

2.4 Comparison of Strengths and Limitations

The backward and forward tracking algorithms, while both designed to elucidate the lifecycle of super-droplets, offer complementary capabilities and are each suited to different scientific questions and computational constraints. Their respective strengths and limitations are summarized below.

The backward tracking algorithm is exceptionally powerful for diagnostic studies that begin from a known end-state. Its primary strength lies in investigating the historical trajectories of specific super-droplets of interest, such as examining the sequence of microphysical processes that lead to the formation of large cloud droplets or raindrops. This makes it an ideal tool for source attribution and for studying the diverse “historical paths” of particles. For example, it can be used to trace complex trajectories like eddy hopping—a process where droplets are detrained from a cloudy updraft, experience subsaturated conditions in the surrounding environment, and are then re-entrained into a cloudy updraft to continue growing (Cooper, 1989; Grabowski and Abade, 2017; Abade et al., 2018). Furthermore, this method provides a potential technical means to investigate the effects of entrainment and mixing. Under extremely inhomogeneous mixing conditions, some cloud

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

droplets evaporate completely while others remain unaffected, potentially growing in the locally supersaturated environment created by the evaporation of their neighbours, which contributes to the broadening of the droplet spectrum (Baker et al., 1980; Burnet and Brenguier, 2007). The backward algorithm allows for tracing the history of these surviving droplets to quantify such effects. A key practical advantage is its efficiency in post-processing, as the direct historical links established at every time step make tracing a lineage computationally fast.

A significant consideration for the backward tracking algorithm is its data storage requirement. Because the algorithm is designed to enable a posteriori selection—allowing researchers to identify and trace any target particle (e.g., a rare "lucky droplet") identified only at the end of the simulation—it currently preserves the lineage pointers for all super-droplets. While this comprehensive logging is essential for diagnosing unpredicted microphysical pathways, it entails a higher storage overhead compared to the forward method. Consequently, applying this full-lineage approach to high-resolution 3D simulations requires careful management of output frequency or variable selection. Future developments may explore adaptive or subset-based backward logging to further reduce this footprint for extremely large-scale applications.

The forward tracking algorithm, by contrast, is designed for prognostic studies and statistical analysis of a representative particle cohort. It allows researchers to randomly sample a specified proportion of super-droplets from the initial stage of the simulation and track their future evolution. Its main advantage is that storage usage is controllable and independent of the total number of super-droplets, making this approach highly appropriate for high-resolution simulations with a large particle count. This makes it particularly well-suited for idealized numerical experiments, such as studies on cloud seeding or weather modification, where the objective is to track the impact and dispersion of a specific set of introduced particles.

The primary drawback of this approach is the a priori nature of the selection. The most "interesting" particles, such as the ones that initiate the first significant precipitation, may not be part of the initially selected cohort, leading to potential missed scientific insights. Additionally, during long-duration simulations, the spatial distribution of the initially selected particles may become highly uneven, which can compromise their statistical representativeness for the domain as a whole. Finally, post-processing the output can be more computationally intensive than for the backward tracking method, as it may involve searching and filtering large datasets to isolate the information for the tagged particles.

3 Application and Results: A Stratocumulus Case Study

3.1 Case Description

The case study employed in this research is the second research flight of the second Dynamics and Chemistry of Marine Stratocumulus field campaign (DYCOMS-II RF02) (Stevens et al., 2003). This case, part of the international DYCOMS project, is based on observations of nocturnal marine stratocumulus off the coast of Southern California on July 11, 2001. It is characterized by a bimodal cloud field, containing regions of closed-cell stratocumulus with light drizzle, interspersed with open-cell areas producing stronger precipitation. The initial thermodynamic conditions represent an average state of these

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

two cloud regimes, while the microphysical conditions are representative of the open-cell regions. Due to its idealized boundary conditions, initial fields, and the availability of rich observational data, the DYCOMS-II RF02 case has been widely used for the evaluation and validation of cloud models, particularly for LES.

3.2 Simulation Setup

3.2.1 General Model Configuration

All simulations were performed using the SCALE-SDM model, which solves the compressible Navier-Stokes equations for moist air using a finite-volume method with a forward-in-time integration scheme. The spatial discretization of Eulerian variables is performed on an Arakawa-C staggered grid (Arakawa and Lamb, 1977). A fourth-order central difference scheme is used for the advection of prognostic dynamical variables, while a third-order upwind scheme by Koren (1993) is used for tracer advection. Other spatial derivatives are approximated using a second-order central difference scheme. For time integration, a fourth-order Runge-Kutta scheme is applied to dynamical variables, and a third-order Runge-Kutta scheme by Wicker and Skamarock (2002) is used for tracers. SGS turbulence is parameterized using a Smagorinsky-Lilly type scheme (Lilly, 1962; Smagorinsky, 1963), which includes stratification effects. To ensure numerical stability, fourth-order hyper-diffusion is added, with the dimensionless coefficient set to 10^{-4} as defined in (Nishizawa et al., 2015).

The SDM explicitly calculates the temporal evolution of aerosol, cloud, and precipitation particles. The microphysical processes considered in this study include advection, sedimentation, evaporation/condensation (including activation/deactivation of cloud condensation nuclei), and, in some simulations, collision-coalescence. To handle the stiffness problem, the condensation/evaporation process is solved using an implicit Euler scheme. The coalescence process is handled by the Monte Carlo stochastic algorithm of Shima et al. (2009). Super-droplets are initialized using a uniform sampling method (see section 5.3 of Shima et al. (2020)). The interaction between super-droplets and SGS turbulence is not directly considered in the current version of the model. For more details on the governing equations and numerical schemes, refer to the works of Shima et al. (2009) and Shima et al. (2020).

The initial vertical profiles for wind, humidity, and temperature follow the setup of Ackerman et al. (2009), with doubly periodic lateral boundary conditions. A simplified radiation model as described in Ackerman et al. (2009) is used. Following the precedent of Ackerman et al. (2009), supersaturation was capped at 1% during the first hour of simulation. They adopted this constraint to prevent excessive activation of Aitken-mode particles during the initial spin-up, which could otherwise hinder precipitation development compared to simulations with fixed droplet number concentrations. The initial aerosol population follows the bimodal lognormal distribution specified in Ackerman et al. (2009). Constant surface sensible and latent heat fluxes are considered, with the latter slightly reduced from 93 W/m^2 to 86.7132 W/m^2 to prevent overestimation of the liquid water content. Momentum exchange between super-droplets and the fluid is also taken into account.

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

3.2.2 Tracking Experiment Setups

To demonstrate the distinct capabilities of the forward and backward tracking algorithms, two different experimental setups were designed. For the forward tracking experiments, designed to investigate the effect of the coalescence process on cloud droplet growth, the simulation domain was set to 2 km × 2 km × 1.5 km with a grid resolution of 20 m × 20 m × 5 m. Two parallel simulations were performed for one hour of spin-up followed by a 10-minute analysis period: one with gravitational coalescence enabled (CTL) and one with it disabled (no_coal), both using a super-droplet number concentration (SDNC) of 40 per cell. In these simulations, 400,000 super-droplets were selected for tracking at the initial time from a total of approximately 7.95 million candidates. The selection criteria were deliberately chosen to focus on super-droplets with a high potential for further condensational or collisional growth: an initial radius greater than or equal to 1 μm and a height within the approximate cloud layer, defined as the range from 50 m below the mean cloud base to 50 m above the mean cloud top (the heights are based on Yin et al. (2024)). This excludes very small aerosols or those far from the cloud (e.g., in the free atmosphere or near the sea surface) that are unlikely to ever activate and are thus not of primary interest for this study. To ensure a representative sample, a stratified proportional random sampling method was used, selecting super-droplets from each 5 m height layer in proportion to the number of candidates within that layer. The exact same set of super-droplets was tracked in both the CTL and no_coal experiments for a rigorous comparison. The output interval of Eulerian fields and super-droplet data are both 5 seconds.

To complement this analysis, a backward tracking experiment, hereafter referred to as BKW_cnd, was conducted to capture the growth histories of potentially significant super-droplets that might have been missed by the initial sampling. This simulation was performed in a quasi-two-dimensional (2D) setup with a coarser resolution to manage the significant computational demands. The current version of the model does not support a strictly 2D simulation, so a thin domain with two grid points was used in the y-direction. A high-resolution 3D simulation for backward tracking would require an immense amount of storage to save the data for all super-droplets at each output step and would also pose a considerable post-processing burden. The quasi-2D approach provides a practical compromise, balancing computational feasibility with the need to capture large-scale eddy structures in the primary x-direction. The domain was configured to 6 km × 0.1 km × 1.5 km with a grid resolution of 50 m × 50 m × 5 m. The simulation ran for one hour of spin-up followed by a 30-minute analysis period, with both Eulerian fields and super-droplet data output every 5 seconds to ensure high temporal continuity. The SDNC was set to 40 per cell. To focus specifically on the contribution of condensational growth, the coalescence process was disabled, and to observe the spatial trajectories more clearly, large-scale horizontal winds were turned off.

The specific settings for these three experiments are detailed in Table 1.

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

3.2.3 Computational Performance

To provide context for the practical application of the developed algorithms, the computational costs associated with the simulations were recorded. All simulations were run on a high-performance computing system equipped with Intel Xeon Gold 6248 (2.5GHz/20C).

The runtime for the forward tracking experiments, which had identical setups apart from the coalescence process, provides insight into the cost of the collision-coalescence calculation itself. The CTL experiment, simulating a total of 70 minutes, required approximately 4,000 core-hours to complete, whereas the no_coal experiment completed in approximately 3,890 core-hours. The backward tracking experiment (BKW_cnd), with its different domain size, resolution, and longer 90-minute simulation period, is not directly comparable in total runtime but required approximately 600 core-hours.

A key trade-off between the two tracking methods is data storage. For the forward tracking simulation (CTL), the primary data consists of 841 time steps of output for the 400,000 selected super-droplets. These files (prefixed with SD_selected_NetCDF_*) contain essential variables including 3D coordinates, terminal velocity, radius, multiplicity, aerosol mass, phase state, and the permanent sd_id, dm_id, and if_coal flags, amounting to approximately 76 MB. The associated coalescence event log (prefixed with SD_coal_output_NetCDF_*), which recorded events involving only these tagged particles at variable time steps, was 48 MB. In contrast, the backward tracking experiment (BKW_cnd) required storing the state of all approximately 2.88 million super-droplets at each of its 1,081 output intervals. These files (prefixed with SD_all_NetCDF_*) contain similar physical variables but store the historical pre_sdid and pre_dmid instead of permanent IDs, resulting in a significantly larger main data volume of approximately 8.5 GB. As coalescence was disabled in this specific experiment, its event log was empty.

The efficiency of the post-processing workflows was also evaluated. The initial data collation step for the forward tracking data, which involved reading and sorting all output files to reconstruct the histories of the 400,000 tracked particles, required roughly 2 hours on a standard workstation. It should be noted that this current workflow utilizes a Python-based script parallelized over 4 CPU cores. Given the interpretive nature of Python, there is significant room for optimization; for instance, reimplementing this sorting-intensive task in a compiled language like Fortran or C++ could drastically reduce the runtime.

Conversely, the backward tracking analysis, implemented as a serial Fortran program, reconstructed the histories of 82,000 seed particles in approximately 20 minutes on the same machine. While a direct performance comparison is complicated by the different implementation languages and particle counts, this result nonetheless highlights the inherent algorithmic advantage of the backward method's $O(1)$ direct lookup, which efficiently retrieves lineage chains from the significantly larger total dataset without the need for global sorting.

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

3.3 Forward Tracking Results: Demonstrating Cohort Analysis

To demonstrate the utility of the forward tracking algorithm, we analyse the growth history of the selected super-droplet cohort to investigate the role of collision-coalescence in cloud droplet evolution. The manageable data output from this method enables a detailed statistical analysis of the 400,000 tracked super-droplets in both the CTL (with coalescence) and no_coal (without coalescence) experiments.

3.3.1 Statistical Evolution of the Tracked Cohort

Figure 2 presents the temporal evolution of the particle size spectrum over the entire 70-minute simulation period for both experiments. The colour shading represents the normalized probability density ($\mathrm{d}P/\mathrm{d}\log r$), providing a comprehensive view of how the droplet population is distributed across different sizes.

In the initial phase of the simulation (within the first ~2 minutes), a high-density spectral band of small super-droplets are rapidly activated and grow by condensation due to the high ambient supersaturation, causing the mean radius (white solid line) to quickly exceed 1 μm. Subsequently, the condensational growth rate slows down, leading to a flattening of the mean radius curve. This trend is consistent with the consumption of available water vapor during the cloud development. After approximately 30 minutes, the mean radius begins a slow decline. This phase is likely associated with the dilution of the cloud layer due to entrainment and mixing, while the overall spectrum of droplet radii gradually broadens, as evidenced by the widening of the probability density shading.

A comparison between the CTL and no_coal experiments (Figure 2) clearly reveals the significant impact of coalescence on droplet growth. In the CTL experiment, the probability density tail extends beyond 20 μm within just 5 minutes of the simulation start. After 30 minutes, the 75th percentile (Q3, upper grey dashed line) is primarily distributed between 15-20 μm, while the 25th percentile (Q1, lower grey dashed line) is concentrated around 1-2 μm. This separation aligns with the two distinct high-density regions visible in the colour shading, indicating a distinct bimodal structure in the droplet population.

In contrast, in the no_coal experiment where droplets grow only by condensation, the spectrum remains narrow. It takes approximately 15 minutes the maximum radius (the upper cyan line) to reach 20 μm. After 30 minutes, the Q3 is mainly in the 10-15 μm range, with Q1 still clustered around 1-2 μm. This demonstrates that without the coalescence process, condensational growth alone is significantly slower and insufficient for the formation of larger droplets (Yau et al., 1979).

The evolution of the droplet size distribution (DSD) at different moments further corroborates these trends (Figure 3). At the beginning of the simulation, the DSDs from both experiments are nearly identical, showing a single-peak distribution concentrated at the small-particle end. As the simulation progresses, the DSDs broaden towards larger sizes and gradually evolve into a bimodal structure, with peaks near 0.7 μm and 10 μm. Around the 30-minute mark, the impact of coalescence becomes evident: a distinct "large-droplet tail" of raindrops with radii exceeding 40 μm appears in the CTL experiment's

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

DSD, a feature entirely absent in the no_coal experiment. This result again underscores the critical importance of the coalescence process for the formation of precipitation.

3.3.2 Statistical Analysis of Coalescence Events

In the current version of SCALE-SDM, droplet growth by coalescence considers only gravitational collection; the effects of turbulence on coalescence are not yet included. To understand the simulation results, we first consider the fundamental physics governing this process. The coalescence probability, P_{jk}, between two real droplets (j, k) within a volume ΔV over a microphysical time step dt_mp is determined by their physical properties (Shima et al., 2009; Shima et al., 2020):

$$P_{jk} = E_{coal}(R_j, R_k)\pi(R_j + R_k)^2 \left|v_j^{\infty} - v_k^{\infty}\right| dt\_mp/\Delta V \qquad (1)$$

where R represents the radius, v^{∞} is the terminal fall velocity, and E_{coal} is the collection efficiency. In this study, the collection efficiency is calculated using the Hall's kernel (Hall, 1980) as modified by Seeβelberg et al. (1996) and Bott (1998). Since each super-droplet represents a multiplicity of real droplets with identical radii and velocities, P_{jk} directly dictates the likelihood of coalescence events in the SDM algorithm. Figure 4 illustrates this theoretical coalescence probability derived from Eq. (1). As expected, P_{jk} increases with the size of the droplets and is highest when the radii of the two colliding droplets are similar.

To investigate the actual occurrence of coalescence events in the simulation, detailed information from the events involving the selected super-droplets was output in the CTL experiment. Figure 5 presents the density of physical coalescence events reconstructed from the Lagrangian tracking logs. By weighting each super-droplet coalescence with its corresponding multiplicity, we recover the actual frequency of coalescence in the cloud. To exclude the collection process by large raindrops, only coalescence events between cloud droplets with radii between 0-20 μm are shown. The analysis reveals that for coalescence to occur, at least one of the participating droplets must grow to a radius of approximately 10 μm. The statistical analysis reveals that coalescence events occur most frequently when droplet radii fall within the 10-20 μm range. This suggests that bridging the gap to reach a radius of approximately 15-20 μm is a critical prerequisite for the efficient initiation of precipitation within the cloud.

The preceding analysis indicates that cloud droplets with a radius surpassing ~15 μm towards 20 μm play a critical role in the stratocumulus precipitation process. While classical theory often cites 20 μm as the threshold for rapid coalescence growth, our results align with observational studies (e.g., Rosenfeld and Gutman (1994)) suggesting that the onset of precipitation initiation begins effectively once droplets exceed the 15 μm threshold. They serve as both the "starting point" for significant coalescence growth and the "target" that condensational growth must achieve. This raises a crucial question: what are the specific environmental conditions and growth pathways that allow some droplets to reach this critical size threshold while others do not? To address this, the following section will utilize the backward tracking method to conduct a detailed analysis of the growth history and environmental conditions of super-droplets that successfully grow beyond 15 μm.

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

3.4 Backward Tracking Results: Demonstrating Lineage Reconstruction

To demonstrate the capability of the backward tracking algorithm, we selected a cohort of super-droplets that grew significantly during the final 30 minutes of the simulation and traced their histories. The primary goal here is to showcase the algorithm's ability to visualize complex particle paths and to link these Lagrangian trajectories to the evolving Eulerian fields of the simulation. The specific selection criterion was: super-droplets with a radius not exceeding 5 μm at the 60-minute mark, but which reached a radius of 15 μm or greater at least once in the subsequent 30 minutes. For clarity, our analysis focuses on the growth phase of these super-droplets, from the 60-minute mark to the moment they reach their maximum radius (t_{max}).

From the super-droplets meeting these criteria, a subset of 1,000 was randomly selected for visualization. Their trajectories are plotted on the Horizontal distance-Height plane in Figure 6. In panel (a), the colour of each trajectory represents the instantaneous radius of the super-droplet, while in panel (b), the colour represents the ambient supersaturation it experienced. The circles mark the initial positions at 60 minutes, and the crosses mark the final positions at t_{max}.

Analysis of these trajectories reveals two key patterns: (1) Vortex-Dominated Motion: The trajectories exhibit distinct spiral-like ascent patterns, indicating that their movement is heavily influenced and controlled by turbulent eddies of various scales. Some trajectories even show a transition from being trapped in small-scale eddies to being transported by larger-scale vortices. (2) Growth Near Cloud Top: The vast majority of these super-droplets reach their maximum radius near the cloud top.

By combining both panels of Figure 6, it becomes evident that the growth of these super-droplets is closely linked to their environmental conditions. Most super-droplets first experience a phase of rapid growth near the cloud base, driven by strong vertical moisture transport and high supersaturation in this region, which facilitates the activation and initial condensational growth of numerous aerosols. As they are transported towards the cloud top by vortices, localized regions of high supersaturation further enhance their condensational growth, allowing them to finally reach the critical size (above 15 μm, entering the coalescence onset range) favourable for initiating precipitation via coalescence.

To provide context for these trajectories, we analysed the corresponding vertical cross-sections at y = 25 m at three different times: 60, 75, and 90 minutes. Figure 7 and Figure 8 show the vertical distribution of turbulent kinetic energy (TKE) and vertical velocity (W), and supersaturation and mean cloud droplet radius, respectively.

At 60 minutes (first row of Figure 7), the boundary layer exhibits well-developed turbulent mixing, with high-TKE regions forming distinct large-eddy ring-like structures, indicating active convection. The vertical velocity field shows an alternating pattern of updrafts and downdrafts, corresponding to multiple vortex structures of similar scales. The regions of high TKE align well with areas of strong vertical velocity. High supersaturation values are primarily found near the cloud base and in regions corresponding to updrafts (positive W) (Figure 8), confirming that vertical moisture transport at the cloud base is a key driver for activation and initial condensational growth. The mean radius generally increases with height and shows local maxima in updraft-dominated regions, indicating effective growth during ascent. The selected super-droplets at this time (see

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

Figure 7) are mostly distributed near the cloud base and in the cloud's middle section, showing good correspondence with the regions of high TKE, updrafts, and high supersaturation.

By the middle of the simulation (90 minutes, second row of Figure 7), the high-TKE regions have intensified with more pronounced vortex structures, indicating enhanced turbulent mixing. The intensity and scale of the vertical velocity are comparable to the 60-minute mark. High supersaturation values remain concentrated near the cloud base but have slightly decreased (Figure 8). Pockets of high mean radius now appear near the cloud top, indicating that droplets continue to grow effectively in the upper parts of the cloud.

In the later stage (90 minutes, third row of Figure 7), the TKE intensity and vertical velocity distribution remain largely unchanged. The overall spatial distribution of supersaturation also shows little change, with some local high-value regions persisting near the cloud top (Figure 8). This suggests that vertical moisture transport remains strong. The mean radius distribution does not show significant changes from the 75-minute mark.

This cross-sectional analysis clearly illustrates the interaction between dynamics and microphysics within the stratocumulus cloud. Vertical transport at the cloud base provides the conditions for droplet activation and initial growth; boundary-layer turbulence facilitates the vertical transport of moisture and droplets, driving condensational growth in updrafts (Yau et al., 1979); and local high-supersaturation regions near the cloud top provide a favourable environment for further growth. It is worth noting that the TKE in a quasi-2D simulation with periodic boundary conditions may be slightly enhanced compared to a full 3D simulation.

4 Discussion

This study has presented the detailed design and implementation of two complementary super-droplet tracking algorithms—forward and backward—and demonstrated their utility in a stratocumulus case study. The discussion below synthesizes the key scientific findings enabled by these new tools and outlines the current limitations and future research directions.

4.1 Synthesis of Findings

The application of the forward and backward tracking algorithms has provided novel insights into the microphysical pathways of cloud droplet growth that are difficult, if not impossible, to obtain with traditional Eulerian or bulk analysis methods.

The forward tracking experiments (CTL vs. no_coal) unequivocally demonstrated the critical role of collision-coalescence in the development of precipitation. By tracking an identical cohort of super-droplets, we quantitatively confirmed that condensational growth alone is insufficient to produce large droplets, while the inclusion of coalescence leads to a rapid broadening of the droplet size distribution and the formation of a distinct large-droplet tail. Furthermore, the statistical analysis of coalescence events pinpointed that droplets reaching the 15-20 μm range act as a key gateway for the efficient

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

initiation of coalescence, acting as a critical threshold that must be surpassed for precipitation to develop. This finding is consistent with a body of previous research, based on both satellite observations and theoretical work, which identified a critical effective radius of around 13–15 μm as the threshold for the onset of warm rain initiation in stratiform clouds (e.g., Rosenfeld and Gutman (1994); Gerber (1996); Freud and Rosenfeld (2012)). Our results, derived directly from process-level statistics within a Lagrangian framework, provide a robust validation of this critical threshold from a new perspective.

This finding motivated the use of the backward tracking algorithm to answer the subsequent question: how do these crucial precipitation embryos (>15 μm) form? The backward analysis revealed a significant pathway. The droplets originating near the cloud base are carried upwards in strong, turbulent updrafts, growing continuously through condensation. This detailed, mechanistic view, linking large-scale turbulent structures directly to the lifecycle of individual particles that become precipitation embryos, would have been impossible to uncover without the lineage-tracing capability of the backward tracking algorithm. This concept of a small subset of particles experiencing exceptionally favourable growth conditions resonates with the "lucky droplet" or "lucky parcel" hypothesis proposed in previous studies (e.g., Cooper (1989); Magaritz et al. (2009); Magaritz-Ronen et al. (2014); Grabowski and Abade (2017)). Our findings therefore provide a concrete, mechanistic validation for the lucky droplet concept within the context of stratocumulus clouds.

Our results are also complementary to those of Chandrakar et al. (2022), who used a similar Lagrangian microphysics scheme to study the macro-scale transition of stratocumulus from closed to open cells. Their work explains when precipitation becomes important for the cloud field by comparing coalescence and eddy turnover timescales, while our work explains how the initial precipitation embryos themselves are formed by identifying specific microphysical pathways. Together, these studies highlight the power of particle-based models in linking micro-scale processes to macro-scale cloud system evolution.

4.2 Limitations and Future Work

While the developed algorithms have proven to be powerful tools, it is important to acknowledge the limitations of the current study and outline avenues for future research. A primary limitation of the backward tracking experiment is its quasi-2D setup, which was a necessary compromise to manage the immense data storage and post-processing demands of a high-resolution 3D simulation. While the quasi-2D domain captures the essential large-eddy structures in the primary direction, the constrained dimensionality may lead to an overestimation of TKE and alter the turbulent mixing characteristics compared to a full 3D environment. Furthermore, the temporal resolution of the reconstructed lineage in the backward analysis is inherently dependent on the model's main output frequency.

Looking forward, overcoming the storage bottleneck to enable high-resolution 3D lineage analysis is a priority. While the current backward algorithm records the history of all super-droplets to ensure no potential "lucky droplet" is missed, this "brute-force" approach may be excessive for basin-scale or global simulations.

A promising avenue for future development is the implementation of "Subset-based Backward Tracking" strategies. This could be achieved through two complementary approaches: 1. Stochastic Sampling: Recording historical pointers for only a

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

randomly selected proportion (e.g., 1% - 10%) of the total population to obtain statistically representative lineages. 2. Physically-Stratified Sampling: Restricting tracking to specific regions of interest, such as exclusively logging particles within the boundary layer. By filtering out free-tropospheric aerosols that are less likely to participate in cloud formation, computational resources can be focused on the most microphysically active sub-domains.

From a statistical perspective, provided the subset is sufficiently large, it will still capture the representative distribution of growth histories—including rare "lucky" events—with high confidence, while linearly reducing the storage footprint. This optimization would shift the computational burden from I/O storage to post-processing, making full 3D lineage reconstruction computationally feasible on next-generation exascale platforms.

The forward tracking method, while efficient in terms of storage, has the inherent limitation that the initially selected cohort of particles may not include the most "interesting" ones that later evolve to become significant precipitation. Moreover, in long-duration simulations, the spatial distribution of the tracked cohort can become statistically unrepresentative of the entire domain.

These limitations motivate several directions for future work. A direct next step would be to apply the backward tracking algorithm to a full 3D simulation, leveraging next-generation high-performance computing resources, to validate and extend the findings from our quasi-2D study. Furthermore, the preliminary analysis in this work has revealed the potential of the backward tracking tool to identify diverse and potentially novel growth pathways for precipitation embryos, which remains a promising avenue for future studies. A detailed scientific investigation into this entrainment-driven formation of "lucky" droplets is a promising avenue for a future study. The current implementation has focused on warm-cloud microphysics. Extending the tracking algorithms to include ice-phase processes—such as tracking the history of riming and aggregation events or identifying the origin of secondary ice particles—represents a significant and valuable future endeavour. The fundamental logic presented here provides a robust foundation for such an expansion. Finally, the rich, process-level data generated by these tracking algorithms can be used to rigorously evaluate and ultimately improve the microphysical parameterization schemes used in larger-scale weather and climate models, contributing to the reduction of uncertainty in cloud and precipitation forecasts.

5 Summary and Conclusion

This paper addresses a methodological gap in the field of cloud microphysics by presenting the detailed design, implementation, and application of two complementary super-droplet tracking algorithms within a large-eddy simulation framework coupled with the Super-Droplet Method (SDM). While Lagrangian particle analysis is a cornerstone of atmospheric science, a formally documented and computationally efficient framework for tracking the complete lifecycle of cloud hydrometeors, particularly through complex processes like coalescence in a parallel computing environment, has been lacking.

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

The primary contribution of this work is the development of this tracking framework, comprising a backward "lineage" tracking algorithm and a forward "tagging" tracking algorithm. The backward algorithm establishes a direct historical link for every super-droplet at each time step, enabling the efficient reconstruction of a particle's complete microphysical history. The forward algorithm employs a stratified random sampling method to select and assign persistent identifiers to a representative cohort of super-droplets, allowing for detailed prognostic analysis with manageable data storage.

The utility of these algorithms was demonstrated in a case study of marine stratocumulus, yielding several key scientific insights. The forward tracking experiments quantitatively confirmed the critical role of collision-coalescence in initiating precipitation, a process that condensational growth alone could not achieve, and identified the 15-20 μm radius range as a key threshold for the onset of efficient coalescence. Building on this finding, the backward tracking analysis provided a detailed, mechanistic view of the classical formation pathway for these crucial droplets, confirming their condensational growth within strong, turbulent cloud-base updrafts. This detailed, mechanistic view, directly linking large-scale turbulent structures to the lifecycle of individual precipitation embryos, is a unique insight made possible by the developed tracking tools.

In conclusion, the tracking algorithms presented here provide the scientific community with a powerful and versatile toolset to investigate the intricate lifecycle of cloud particles with unprecedented detail. This work not only enables a deeper understanding of fundamental processes like precipitation initiation and entrainment-mixing but also offers a robust methodology for evaluating and improving microphysical parameterizations in larger-scale weather and climate models, ultimately contributing to the reduction of uncertainty in cloud and precipitation simulations.

Code and data availability.

The source code for the algorithms presented in this study is based on and integrated into the SCALE-SDM version 5.2.6-2.3.1 framework. The implementations of the two distinct tracking algorithms are permanently archived and publicly available on Zenodo under separate DOIs. The backward tracking algorithm code is available at: https://doi.org/10.5281/zenodo.15845346 (Yin, 2025b) and the forward tracking algorithm code is available at: https://doi.org/10.5281/zenodo.15845351 (Yin, 2025a). The development version of the code, as well as the post-processing scripts used to generate the results in this study and detailed usage instructions in the Readme.md file, are available in the public GitHub repository: https://github.com/wangyouyue/SDM_product-230225_Yin-SD_tracking (last access: 11 Dec 2025). All the model results used for this study can be reproduced by following the instructions included in the above repository. Due to the size of the model results, the data are deposited in local storage at the University of Hyogo in Kobe, Japan, and are available from the corresponding author upon request.

Video supplement.

The video supplement related to this article is available online at https://doi.org/10.5281/zenodo.15878178 (Yin, 2025c).

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

Author contribution.

All the authors designed the experiments and CY carried them out. SS and CY developed the model code and CY performed the simulations. CY prepared the manuscript with contributions from all co-authors.

Competing interests.

The authors declare that they have no conflict of interest.

Acknowledgements.

CY and SS would like to thank Mikito Toda for their generous support and informative discussions. This research partly used the computational resources of Hokkaido University and the University of Osaka through the HPCI System Research Project (project IDs: hp200078, hp210059, hp220062, hp230166, hp240151, hp250136), and the computer facilities of the Center for Cooperative Work on Data science and Computational science, University of Hyogo. We acknowledge the High Performance Computing Center of Nanjing Information Science & Technology for their support of this work. The SCALE library was developed by Team-SCALE of RIKEN Center for Computational Sciences (https://scale.riken.jp/, last access: 11 Dec 2025). Generative AI tools (Gemini) were used to assist with language editing for this manuscript.

Financial support

This research has been supported by the National Natural Science Foundation of China (42325503), JSPS KAKENHI (grant nos. 20H00225, 23H00149, 23K03265), and JST [Moonshot R&D][Grant Number JPMJMS2286, JPMJMS2283]. The appointment of Chunsong Lu at Nanjing University of Information Science & Technology is partially supported by the Jiangsu Specially-Appointed Professor (Grant No. R2024T01).

References

Abade, G. C., Grabowski, W. W., and Pawlowska, H.: Broadening of cloud droplet spectra through eddy hopping: Turbulent entraining parcel simulations, Journal of the Atmospheric Sciences, 75, 3365-3379, https://doi.org/10.1175/JAS-D-18-0078.1, 2018.

Ackerman, A. S., VanZanten, M. C., Stevens, B., Savic-Jovcic, V., Bretherton, C. S., Chlond, A., Golaz, J.-C., Jiang, H., Khairoutdinov, M., and Krueger, S. K.: Large-eddy simulations of a drizzling, stratocumulus-topped marine boundary layer, Monthly Weather Review, 137, 1083-1110, https://doi.org/10.1175/2008MWR2582.1, 2009.

Arakawa, A. and Lamb, V. R.: Computational design of the basic dynamical processes of the UCLA general circulation model, General circulation models of the atmosphere, Supplement C, 173-265 pp.1977.

Ayala, O., Rosa, B., and Wang, L.-P.: Effects of turbulence on the geometric collision rate of sedimenting droplets. Part 2. Theory and parameterization, New Journal of Physics, 10, 075016, 2008a.

Ayala, O., Rosa, B., Wang, L.-P., and Grabowski, W. W.: Effects of turbulence on the geometric collision rate of sedimenting droplets. Part 1. Results from direct numerical simulation, New Journal of Physics, 10, 075015, 2008b.

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

Baker, M., Corbin, R., and Latham, J.: The influence of entrainment on the evolution of cloud droplet spectra: I. A model of inhomogeneous mixing, Quarterly Journal of the Royal Meteorological Society, 106, 581-598, 1980.
Bott, A.: A Flux Method for the Numerical Solution of the Stochastic Collection Equation, Journal of the Atmospheric Sciences, 55, 2284-2293, https://doi.org/10.1175/1520-0469(1998)055<2284:AFMFTN>2.0.CO;2, 1998.
Burnet, F. and Brenguier, J.-L.: Observational Study of the Entrainment-Mixing Process in Warm Convective Clouds, Journal of the Atmospheric Sciences, 64, 1995-2011, https://doi.org/10.1175/JAS3928.1, 2007.
Chandrakar, K. K., Morrison, H., and Witte, M.: Evolution of Droplet Size Distributions During the Transition of an Ultraclean Stratocumulus Cloud System to Open Cell Structure: An LES Investigation Using Lagrangian Microphysics, Geophysical Research Letters, 49, e2022GL100511, https://doi.org/10.1029/2022GL100511, 2022.
Chandrakar, K. K., Morrison, H., Harrington, J. Y., Pokrifka, G., and Magee, N.: What Controls Crystal Diversity and Microphysical Variability in Cirrus Clouds?, Geophysical Research Letters, 51, e2024GL108493, https://doi.org/10.1029/2024GL108493, 2024.
Cooper, W. A.: Effects of Variable Droplet Growth Histories on Droplet Size Distributions. Part I: Theory, Journal of Atmospheric Sciences, 46, 1301-1311, 1989.
Draxler, R. and Hess, G.: An overview of the HYSPLIT_4 modeling system for trajectories, dispersion, and deposition, Australian Meteorological Magazine, 47, 295-308, 1998.
Freire, J., Koop, D., Santos, E., and Silva, C. T.: Provenance for Computational Tasks: A Survey, Computing in Science & Engineering, 10, 11-21, 10.1109/MCSE.2008.79, 2008.
Freud, E. and Rosenfeld, D.: Linear relation between convective cloud drop number concentration and depth for rain initiation, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2011JD016457, 2012.
Gerber, H.: Microphysics of Marine Stratocumulus Clouds with Two Drizzle Modes, Journal of Atmospheric Sciences, 53, 1649-1662, https://doi.org/10.1175/1520-0469(1996)053<1649:MOMSCW>2.0.CO;2, 1996.
Grabowski, W. W. and Abade, G. C.: Broadening of cloud droplet spectra through eddy hopping: Turbulent adiabatic parcel simulations, Journal of the Atmospheric Sciences, 74, 1485-1493, https://doi.org/10.1175/JAS-D-17-0043.1, 2017.
Hall, W. D.: A Detailed Microphysical Model Within a Two-Dimensional Dynamic Framework: Model Description and Preliminary Results, Journal of Atmospheric Sciences, 37, 2486-2507, https://doi.org/10.1175/1520-0469(1980)037<2486:ADMMWA>2.0.CO;2, 1980.
Khain, A., Ovtchinnikov, M., Pinsky, M., Pokrovsky, A., and Krugliak, H.: Notes on the state-of-the-art numerical modeling of cloud microphysics, Atmospheric Research, 55, 0-224, 2000.
Khain, A., Beheng, K., Heymsfield, A., Korolev, A., Krichak, S., Levin, Z., Pinsky, M., Phillips, V., Prabhakaran, T., and Teller, A.: Representation of microphysical processes in cloud-resolving models: Spectral (bin) microphysics versus bulk parameterization, Reviews of Geophysics, 53, 247-322, https://doi.org/10.1002/2014RG000468, 2015.
Koren, B.: A robust upwind discretization method for advection, diffusion and source terms, Centrum voor Wiskunde en Informatica Amsterdam1993.
Krueger, S. K., Su, C.-W., and McMurtry, P. A.: Modeling Entrainment and Finescale Mixing in Cumulus Clouds, Journal of the Atmospheric Sciences, 54, 2697-2712, https://doi.org/10.1175/1520-0469(1997)054<2697:MEAFMI>2.0.CO;2, 1997.
Lilly, D. K.: On the numerical simulation of buoyant convection, Tellus, 14, 148-172, https://doi.org/10.1111/j.2153-3490.1962.tb00128.x, 1962.
Magaritz, L., Pinsky, M., Krasnov, O., and Khain, A.: Investigation of Droplet Size Distributions and Drizzle Formation Using A New Trajectory Ensemble Model. Part II: Lucky Parcels, Journal of the Atmospheric Sciences, 66, 781-805, https://doi.org/10.1175/2008JAS2789.1, 2009.
Magaritz-Ronen, L., Pinsky, M., and Khain, A.: Effects of Turbulent Mixing on the Structure and Macroscopic Properties of Stratocumulus Clouds Demonstrated by a Lagrangian Trajectory Model, Journal of the Atmospheric Sciences, 71, 1843-1862, https://doi.org/10.1175/JAS-D-12-0339.1, 2014.
Matsushima, T., Nishizawa, S., and Shima, S.: Optimization and sophistication of the super-droplet method for ultrahigh resolution cloud simulations, Geosci. Model Dev. Discuss., 2023, 1-53, 10.5194/gmd-2023-26, 2023.
Nishizawa, S., Yashiro, H., Sato, Y., Miyamoto, Y., and Tomita, H.: Influence of grid aspect ratio on planetary boundary layer turbulence in large-eddy simulations, Geoscientific Model Development, 8, 3393-3419, https://doi.org/10.5194/gmd-8-3393-2015, 2015.

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

Pisso, I., Sollum, E., Grythe, H., Kristiansen, N. I., Cassiani, M., Eckhardt, S., Arnold, D., Morton, D., Thompson, R. L., Groot Zwaaftink, C. D., Evangeliou, N., Sodemann, H., Haimberger, L., Henne, S., Brunner, D., Burkhart, J. F., Fouilloux, A., Brioude, J., Philipp, A., Seibert, P., and Stohl, A.: The Lagrangian particle dispersion model FLEXPART version 10.4, Geosci. Model Dev., 12, 4955-4997, 10.5194/gmd-12-4955-2019, 2019.

Rosenfeld, D. and Gutman, G.: Retrieving microphysical properties near the tops of potential rain clouds by multispectral analysis of AVHRR data, Atmospheric Research, 34, 259-283, https://doi.org/10.1016/0169-8095(94)90096-5, 1994.

Sato, Y., Nishizawa, S., Yashiro, H., Miyamoto, Y., Kajikawa, Y., and Tomita, H.: Impacts of cloud microphysics on trade wind cumulus: which cloud microphysics processes contribute to the diversity in a large eddy simulation?, Progress in Earth & Planetary Science, 2, 23, 2015.

Seeβelberg, M., Trautmann, T., and Thorn, M.: Stochastic simulations as a benchmark for mathematical methods solving the coalescence equation, Atmospheric Research, 40, 33-48, https://doi.org/10.1016/0169-8095(95)00024-0, 1996.

Shima, S., Sato, Y., Hashimoto, A., and Misumi, R.: 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, Geosci. Model Dev., 13, 4107-4157, 10.5194/gmd-13-4107-2020, 2020.

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

Simmhan, Y. L., Plale, B., and Gannon, D.: A survey of data provenance in e-science, SIGMOD Rec., 34, 31–36, 10.1145/1084805.1084812, 2005.

Smagorinsky, J.: General circulation experiments with the primitive equations: I. The basic experiment, Monthly weather review, 91, 99-164, https://doi.org/10.1175/1520-0493(1963)091%3C0099:GCEWTP%3E2.3.CO;2, 1963.

Squires, P.: The microstructure and colloidal stability of warm clouds: Part I—The relation between structure and stability, Tellus, 10, 256-261, 1958.

Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., and Ngan, F.: NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System, Bulletin of the American Meteorological Society, 96, 2059-2077, https://doi.org/10.1175/BAMS-D-14-00110.1, 2015.

Stevens, B., Lenschow, D. H., Vali, G., Gerber, H., Bandy, A., Blomquist, B., Brenguier, J.-L., Bretherton, C., Burnet, F., and Campos, T.: Dynamics and chemistry of marine stratocumulus—DYCOMS-II, Bulletin of the American Meteorological Society, 84, 579-594, 2003.

Stohl, A., Forster, C., Frank, A., Seibert, P., and Wotawa, G.: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, Atmos. Chem. Phys., 5, 2461-2474, 10.5194/acp-5-2461-2005, 2005.

Wang, L.-P. and Maxey, M. R.: Settling velocity and concentration distribution of heavy particles in homogeneous isotropic turbulence, Journal of Fluid Mechanics, 256, 27-68, 10.1017/S0022112093002708, 1993.

Wicker, L. J. and Skamarock, W. C.: Time-splitting methods for elastic models using forward time schemes, Monthly weather review, 130, 2088-2097, https://doi.org/10.1175/1520-0493(2002)130%3C2088:TSMFEM%3E2.0.CO;2, 2002.

Yau, K, M., Rogers, and R, R.: A SHORT COURSE IN CLOUD PHYSICS, Pergamon Press1979.

Yin, C.: SCALE-SDM with Forward Tracking (v1.1.1), Zenodo [code], 10.5281/zenodo.15845351, 2025a.

Yin, C.: SCALE-SDM with Backward Tracking (v1.1.1), Zenodo [code], 10.5281/zenodo.15845346, 2025b.

Yin, C.: Video supplement for "A Lagrangian Particle Tracking Framework for the Super-Droplet Method: Development, Implementation, and Application of Backward and Forward Algorithms in SCALE-SDM 5.2.6-2.3.1", 10.5281/zenodo.15878178, 2025c.

Yin, C., Shima, S., Xue, L., and Lu, C.: 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, Geosci. Model Dev., 17, 5167-5189, 10.5194/gmd-17-5167-2024, 2024.

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

```
// --- Main Program Entry ---
Procedure MAIN()
    // 1. Initialize all tracking variables at the beginning of the simulation.
    CALL INITIALIZE_SD_BACKWARD_TRACKING()

    // 2. Loop through each model time step.
    FOR time_step = 1 TO max_time_steps DO
        // Encapsulate all actions for one time step to ensure correct logical order.
        CALL ADVANCE_ONE_TIME_STEP_BACKWARD(time_step)
    END FOR
END PROCEDURE

// --- Initialization Procedure ---
Procedure INITIALIZE_SD_BACKWARD_TRACKING()
    // Set all historical identifiers to an invalid state at t=0.
    FOR each super-droplet i DO
        pre_sdid[i] ← INVALID_ID
        pre_dmid[i] ← INVALID_ID
        if_coal[i] ← 0
    END FOR
END PROCEDURE

// --- Single Time Step Advancement Procedure ---
// This procedure shows the crucial execution order of operations.
Procedure ADVANCE_ONE_TIME_STEP_BACKWARD(time_step)
    // --- Phase 1: Calculate Physics ---
    // Physical evolution is calculated first, using the historical data from the previous step.

    // 1a. Update spatial coordinates of all super-droplets.
    CALL SD_MOTION_TRACKING()

    // 1b. Handle particles crossing MPI domain boundaries.
    //     Passes historical tracking variables (pre_sdid, pre_dmid, if_coal).
    CALL SD_BOUNDARY_HANDLING()

    // 1c. Calculate coalescence and record parent lineage.
    //     This procedure READS from pre_sdid/pre_dmid to identify parents.
    //     It MODIFIES if_coal for droplets that collided in this step.
    CALL SD_COALESCENCE_TRACKING()

    // --- Phase 2: Output Data ---
    // Output is generated after the state for the current time step has been fully calculated.

    IF (is_output_time_step(time_step)) THEN
        // Writes the historical identifiers (pre_sdid, pre_dmid) and current coalescence status
(if_coal) to file.
        // The detailed implementation is in the Appendix.
        CALL OUTPUT_SD_STATE_DATA()
    END IF

    // --- Phase 3: Update History for Next Step ---
    // This is the core of the backward tracking logic, executed at the end of the time step.

    CALL UPDATE_HISTORICAL_IDS()
END PROCEDURE

// --- Historical ID Update Procedure (The Core of the Algorithm) ---
// This procedure explicitly shows the update logic that enables backward tracking.
Procedure UPDATE_HISTORICAL_IDS()
    // For each super-droplet, its current ID becomes the "previous ID" for the next time step.
    FOR each super-droplet i DO
        // 'current_sdid[i]' is the definitive ID of the droplet at the end of the current time step.
        pre_sdid[i] ← current_sdid[i]
        pre_dmid[i] ← current_dmid[i]

        // Reset the coalescence flag for the next tracking interval.
        if_coal[i] ← 0
    END FOR
END PROCEDURE
```

Algorithm 1: A high-level representation of the Backward Tracking algorithm, illustrating the main operational procedures and their execution order within a single model time step. This structure highlights the "use-first, then-update" paradigm critical for maintaining the lineage chain. The full implementation of the algorithm is publicly available in the archived source code (Yin, 2025b).

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

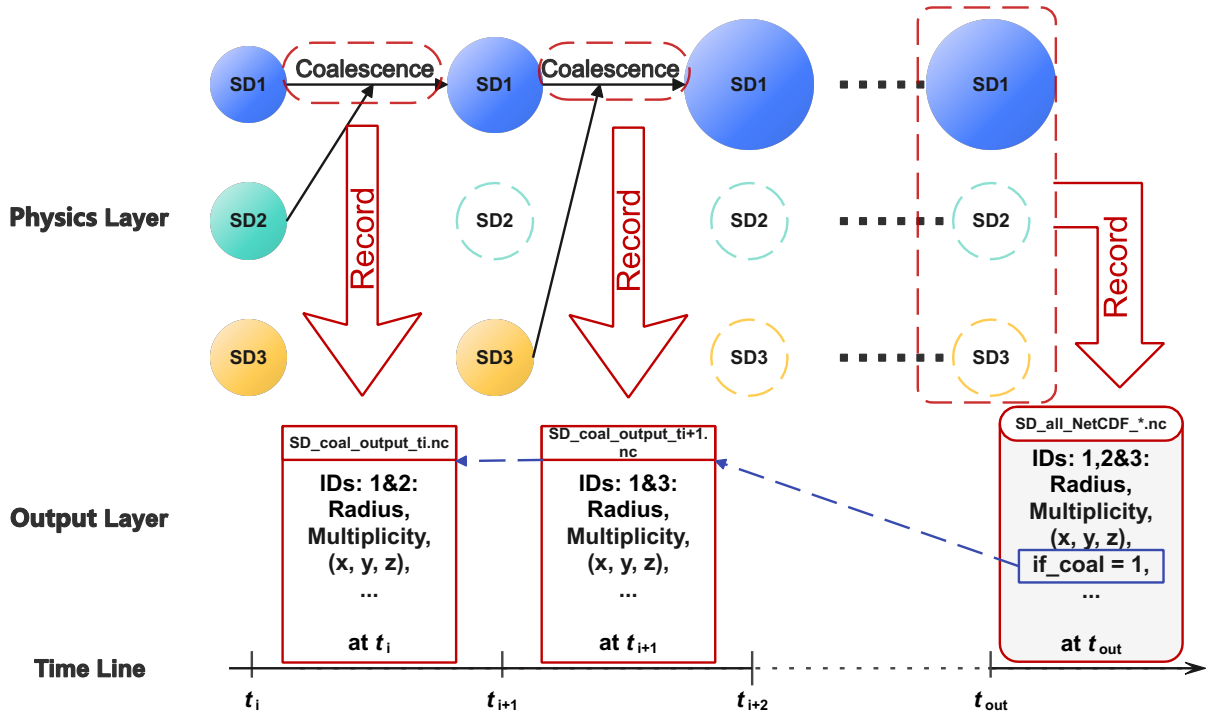


Figure 1 Schematic illustration of the lineage recording mechanism within the backward tracking algorithm. The diagram contrasts the high-frequency microphysical time steps (dt_mp) with the lower-frequency main output interval (dt_out). (Top) The timeline of physical evolution shows a tracked super-droplet (SD1) growing through two distinct coalescence events. Note that the partner super-droplets (SD2, SD3) are updated (e.g., reduced multiplicity) and continue to exist in the simulation (indicated by dashed lines). (Bottom) Each coalescence event is instantly recorded in a dedicated event file (e.g., SD_coal_output_ti.nc and SD_coal_output_ti+1.nc) at the exact microphysical step of occurrence. At the main output time (t_{out}), the state of SD1, SD2 and SD3 are saved in the main output file (SD_all_NetCDF_*.nc) with the if_coal flag set to 1. The blue dashed arrow indicates the backward diagnostic logic: the flag serves as a trigger for the post-processing algorithm to look back into the event logs and reconstruct the complete branching history.

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

```
// --- Main Program Entry ---
Procedure MAIN()
    // 1. Initialize and select a subset of super-droplets for tracking at the beginning of the
simulation.
    CALL INITIALIZE_SD_FORWARD_TRACKING()

    // 2. Loop through each model time step.
    FOR time_step = 1 TO max_time_steps DO
        // Encapsulate all actions for one time step to ensure correct logical order.
        CALL ADVANCE_ONE_TIME_STEP_FORWARD(time_step)
    END FOR
END PROCEDURE

// --- Initialization Procedure ---
// This procedure has a distinct two-stage logic for forward tracking.
Procedure INITIALIZE_SD_FORWARD_TRACKING()
    // Stage 1: Initialize all tracking identifiers to an invalid state.
    FOR each super-droplet i DO
        sd_id[i] ← INVALID_ID
        dm_id[i] ← INVALID_ID
        if_coal[i] ← 0
    END FOR

    // Stage 2: Select a subset of super-droplets based on physical criteria and assign permanent IDs.
    // The detailed implementation of the stratified random sampling is in the Appendix.
    CALL SELECT_AND_TAG_PARTICLES()
END PROCEDURE

// --- Single Time Step Advancement Procedure ---
// This procedure shows the crucial execution order of operations for forward tracking.
Procedure ADVANCE_ONE_TIME_STEP_FORWARD(time_step)
    // --- Phase 1: Calculate Physics for ALL Super-Droplets ---
    // Physical evolution is calculated for the entire ensemble of particles.

    // 1a. Update spatial coordinates.
    CALL SD_MOTION_TRACKING()

    // 1b. Apply stochastic perturbations to particle positions.
    CALL ADD_RANDOM_PERTURBATION_ALL()

    // 1c. Handle particles crossing MPI domain boundaries.
    //     Passes the persistent tracking variables (sd_id, dm_id, if_coal) for tagged particles.
    CALL SD_BOUNDARY_HANDLING()

    // 1d. Calculate coalescence for all particles.
    //     This procedure checks if any tagged particles are involved in collisions and records their
fate,
    //     including marking them as 'lost' if consumed.
    CALL SD_COALESCENCE_TRACKING()

    // --- Phase 2: Output Data for Tracked Particles ---
    // Output is generated after the state for the current time step has been fully calculated.

    IF (is_output_time_step(time_step)) THEN
        // Writes the state data (position, radius, etc.) and tracking identifiers
        // ONLY for the super-droplets that are currently being tracked (sd_id != INVALID_ID).
        CALL OUTPUT_TRACKED_SD_DATA()

        // After outputting, reset flags for the next tracking interval.
        CALL RESET_COALESCENCE_FLAGS()
    END IF
END PROCEDURE

// --- Data Output and Flag Reset Procedure ---
// This procedure handles the output of tracked particles and prepares for the next interval.
Procedure OUTPUT_TRACKED_SD_DATA()
    // Implementation details are in the Appendix. In summary:
    // FOR each super-droplet i DO
    //     IF (sd_id[i] != INVALID_VALUE) THEN
    //         WRITE_TO_OUTPUT_FILE(particle_i_data...)
    //     END IF
    // END FOR
END PROCEDURE

Procedure RESET_COALESCENCE_FLAGS()
    // For each super-droplet, reset the coalescence flag for the next output interval.
    // The permanent sd_id and dm_id are NOT changed.
    FOR each super-droplet i DO
        if_coal[i] ← 0
    END FOR
END PROCEDURE
```

Algorithm 2 A high-level representation of the Super-Droplet Forward Tracking algorithm. The algorithm begins by selecting and tagging a subset of super-droplets and then follows their evolution through the model's physical processes at each time step. The full implementation of the algorithm is publicly available in the archived source code (Yin, 2025a).

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

Table 1 Summary of experimental setups for the forward and backward tracking simulations. dt_mp and dt_dyn represent the microphysical time and dynamical time step, respectively. In the microphysical setup (MP setup), cnd represents the condensation process and coal represents the coalescence process.

Experiments	Tracking Type	Domain (km^3)	Resolution (m^3)	Time Steps (dt_mp, dt_dyn) (s)	SDNC (#/cell)	MP setup
CTL	Forward	2 × 2 × 1.5	20 × 20 × 5	0.1, 0.01	40	cnd + coal
no_coal	Forward	2 × 2 × 1.5	20 × 20 × 5	0.1, 0.01	40	cnd
BKW_cnd	Backward	6 × 0.1 × 1.5	50 × 50 × 5	0.2, 0.02	40	cnd + coal

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

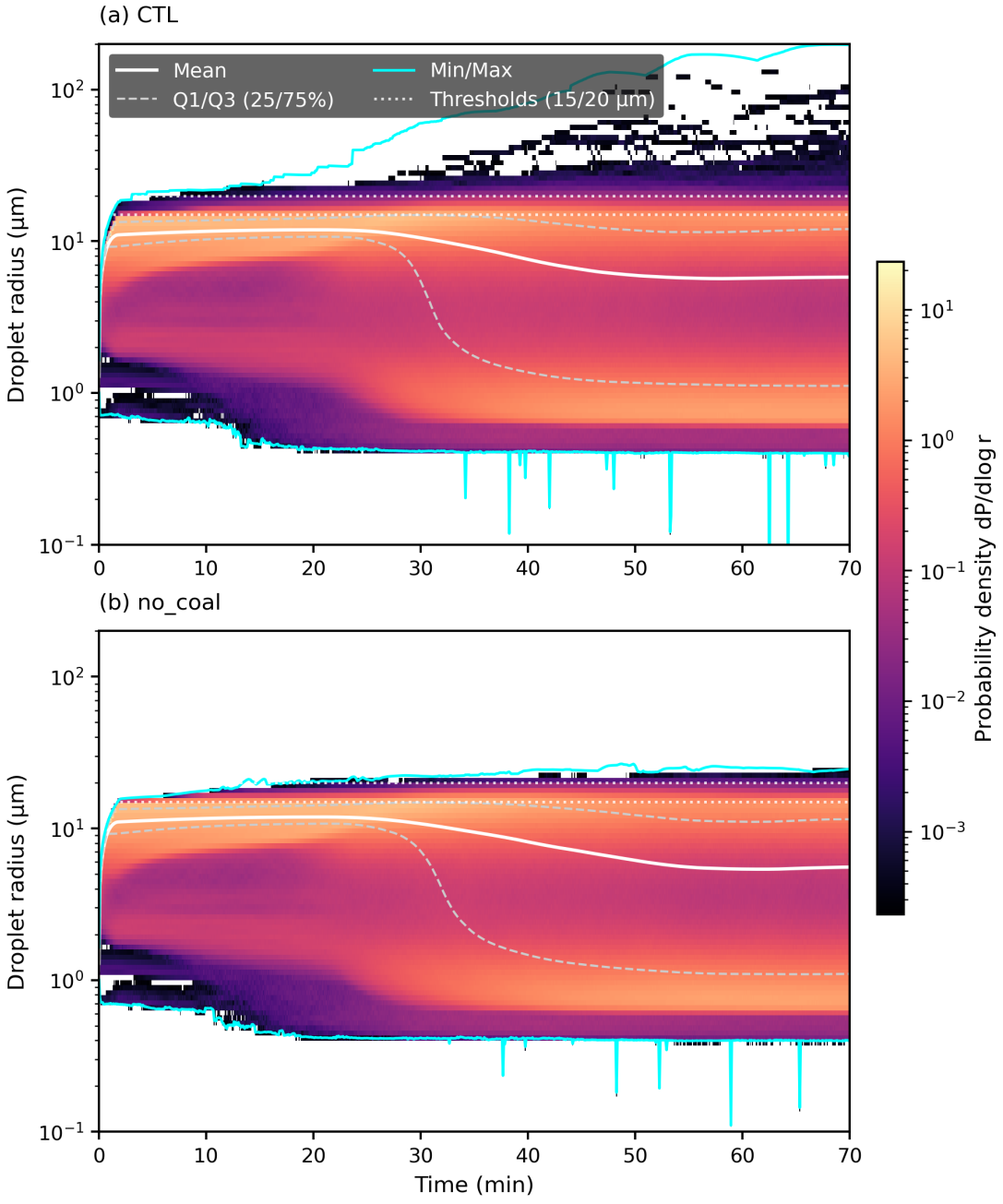
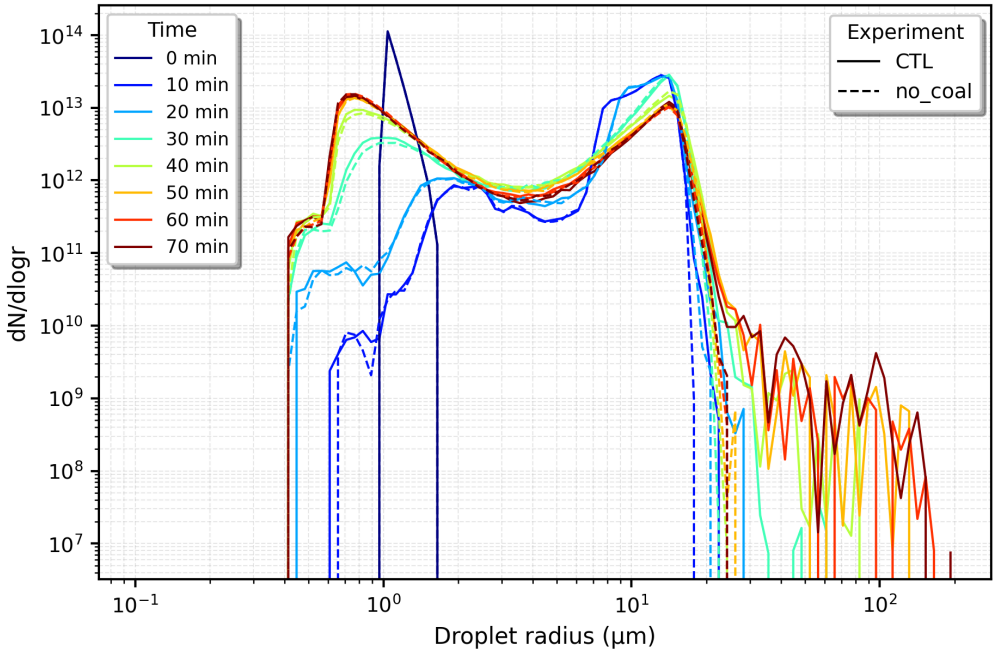


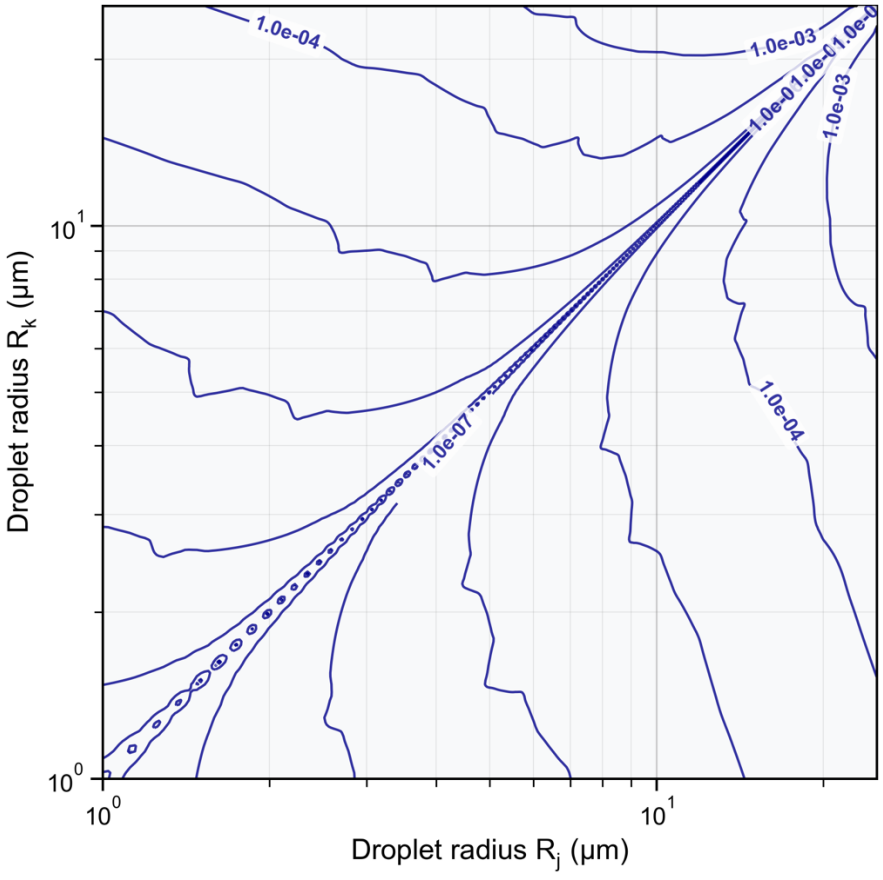
Figure 2 Time evolution of the particle size spectrum in the (a) CTL and (b) no_coal experiments. The colour shading represents the normalized probability density distribution ($\mathrm{d}P/\mathrm{d}\log r$) on a logarithmic scale, illustrating the broadening of the droplet spectrum over time. Overlaid lines indicate the statistical properties of the droplet population: the mean radius (white solid line), the minimum and maximum radii (cyan solid lines), and the 25th and 75th percentiles (grey dashed lines). The horizontal white dotted lines mark the 15 μm and 20 μm radius thresholds, serving as references for the onset of efficient coalescence.

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025



800 **Figure 3 Evolution of cloud droplet size distribution (DSD) in the CTL (solid lines) and no_coal (dashed lines) experiments. The vertical axis represents the DSD ($\mathrm{d}N/\mathrm{d}\log r$) of the population of tracked particle, where N is the number of tracked real droplets, not a number concentration. Different coloured lines represent the DSD at different times (from 0 min to 70 min of the simulation, with an interval of 10 min).**

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025



805

Figure 4 Contour plot of coalescence probability (P_{jk}) for real droplet pairs (j, k). The horizontal and vertical axes represent the radii R_j and R_k (μm) of the droplet pair, respectively. Only results for radii ranging from 1 μm to 25 μm are shown.

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

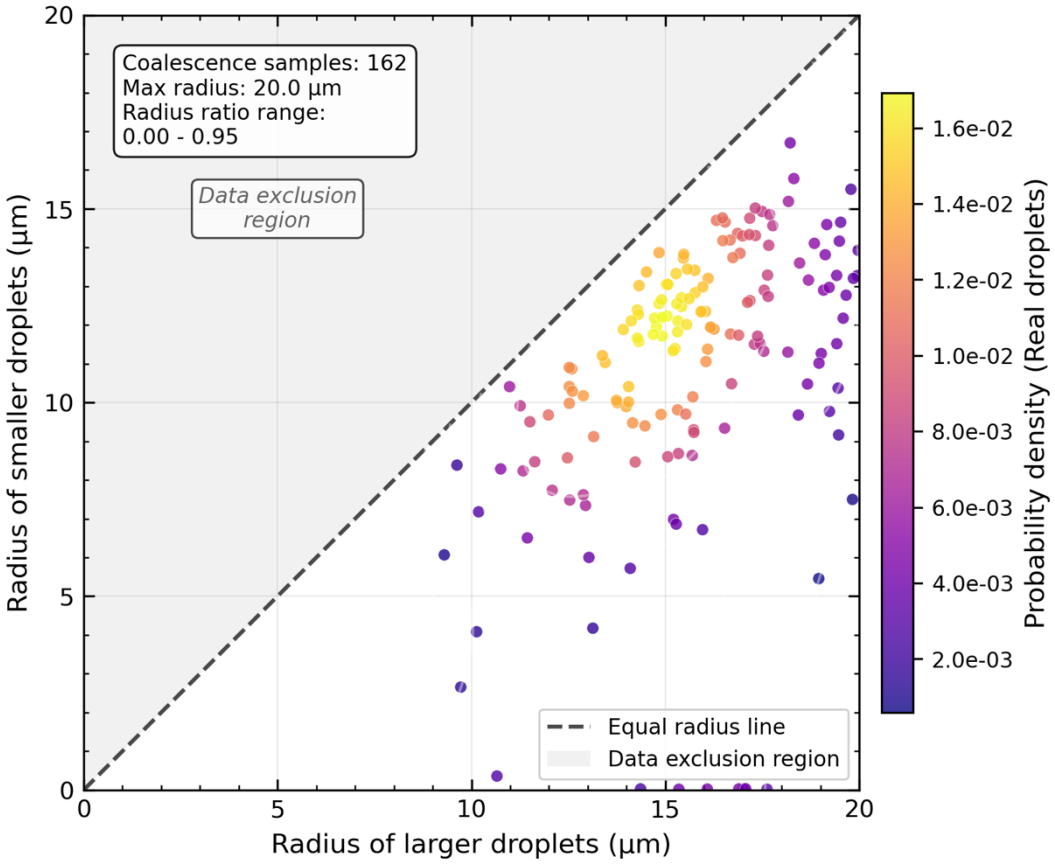


Figure 5 Scatter plot of realized coalescence events weighted by super-droplet multiplicity in the CTL experiment. The horizontal and vertical axes represent the larger and smaller radii (µm) of the coalescing super-droplet pairs, respectively. The colour shading represents the event density of real droplet coalescence (estimated via Gaussian kernel density estimation), highlighting the specific radius range (10-20 µm) where collisions are most concentrated, as opposed to the theoretical probability shown in Figure 3. The analysis is strictly confined to the initiation phase; thus, only collision events between droplets with radii smaller than 20 µm are included to isolate the warm rain initiation bottleneck from the subsequent rapid growth of large raindrops.

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

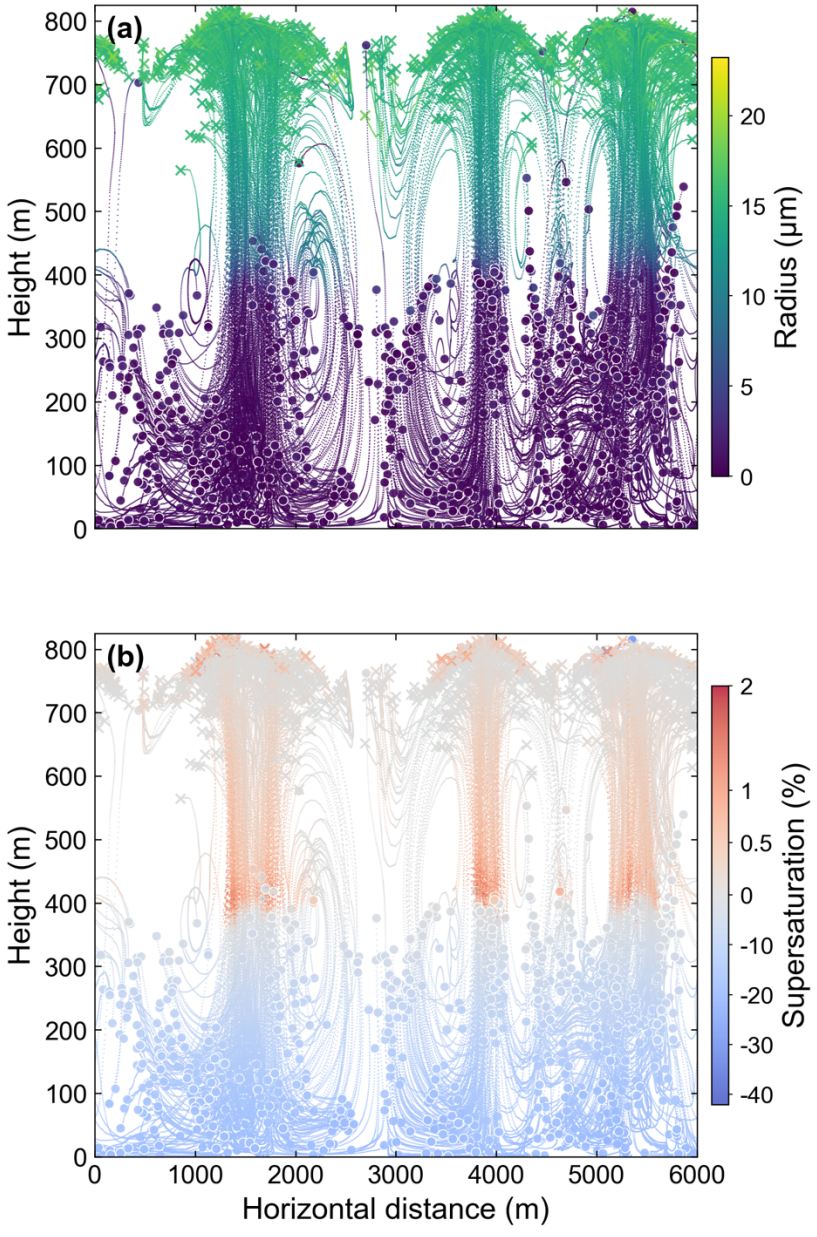


Figure 6 Spatial trajectories of 1, 000 randomly selected super-droplets that meet the large droplet criteria in the Horizontal distance (m)-Height (m) plane. (a) Trajectory colours represent the instantaneous radius of the super-droplet at the corresponding location; (b) Trajectory colours represent the ambient supersaturation at the corresponding location. Circles indicate the initial positions of the super-droplets at the 60th minute, and crosses indicate their positions when they reach their maximum radii.

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

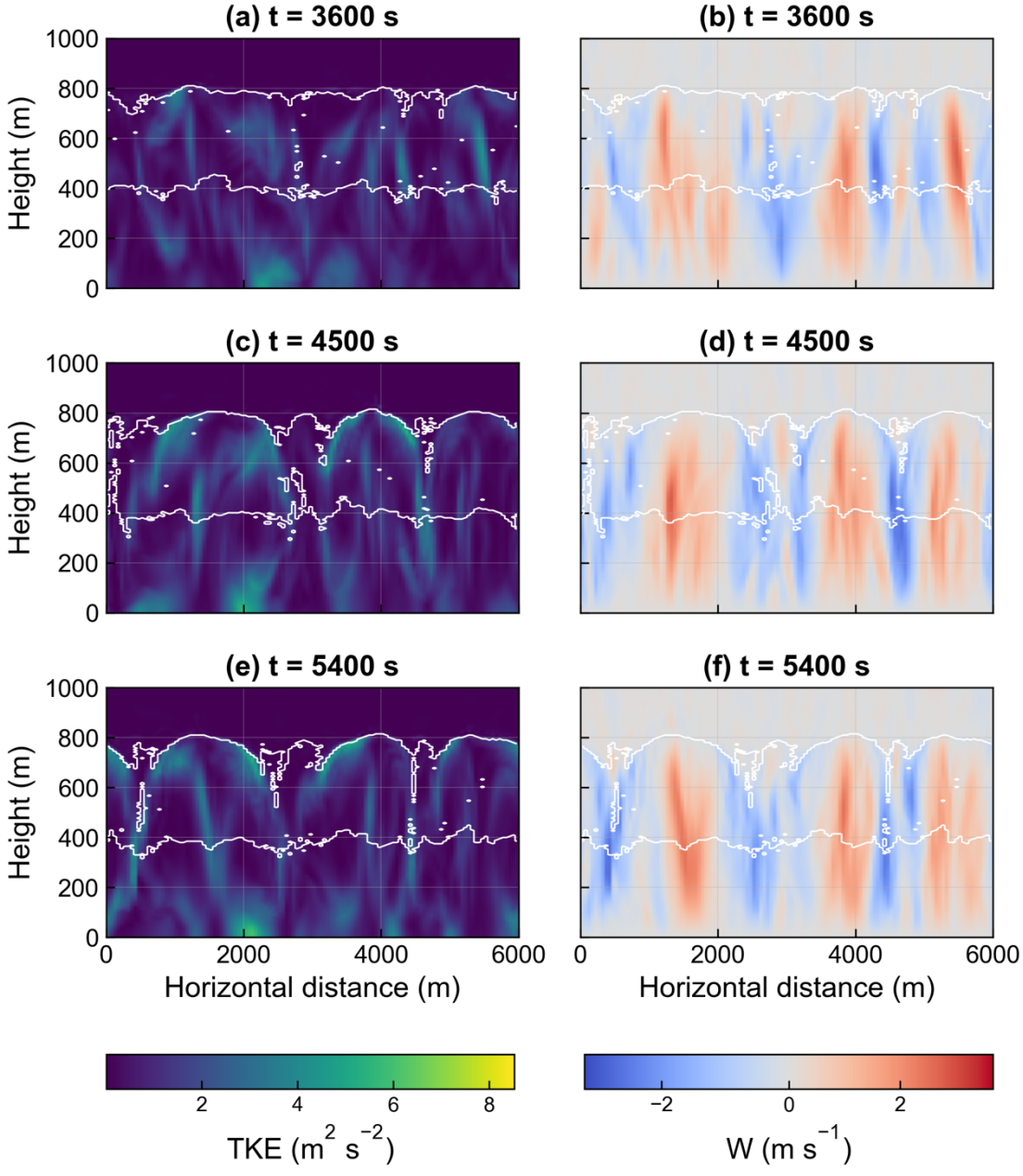


Figure 7 Vertical cross-sections of turbulent kinetic energy (TKE, m²/s², left column) and vertical velocity (W, m/s, right column) at y = 25 m (representing the cross-section of the quasi-2D domain) for simulation times of t = 3600 s (60 min) (first row), t = 4500 s (75 min) (second row), and t = 5400 s (90 min) (third row). White contours represent the cloud boundaries.

https://doi.org/10.5194/egusphere-2025-6221
Preprint. Discussion started: 18 December 2025

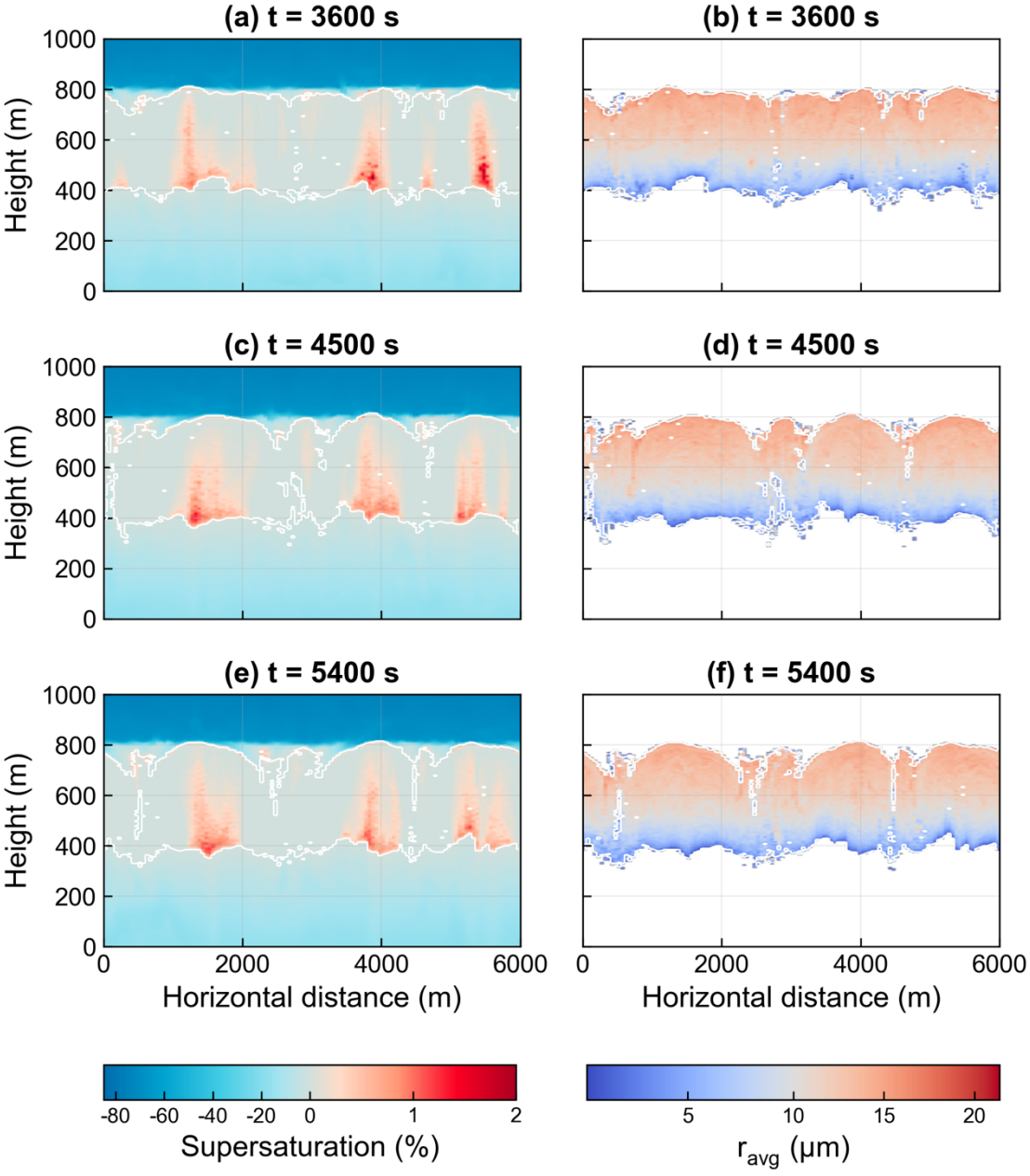


Figure 8 Vertical cross-sections of supersaturation (%, left column) and mean cloud droplet radius (μm, right column) at y = 25 m (representing the cross-section of the quasi-2D domain) for simulation times of t = 3600 s (60 min) (first row), t = 4500 s (75 min) (second row), and t = 5400 s (90 min) (third row). White contours represent the cloud boundaries.