

## **Response to the Topic Editor and Referees**

We thank the Topic Editor and both referees for their careful assessment of the revised manuscript. We are encouraged that the editor has classified the remaining work as a minor revision and that both reports recognise the substantial improvement of the manuscript. We have addressed all remaining comments. The principal changes are: (1) simplifying the title and abstract; (2) adding normalised computational metrics with explicit denominators; (3) clarifying the scope of the quasi-two-dimensional fraction tests and the fixed-5 % SDNC scaling experiment; (4) distinguishing target-ID transfer from code-level coalescence-history verification; (5) adding a complete comparison of the recorded coalescence histories from the two TPHT runs; and (6) documenting the computational conditions required for reproducibility. We also corrected the core-hour accounting used in archived auxiliary analyses after identifying that a scheduler-inherited OpenMP setting did not represent the non-OpenMP executable. All changes are shown in the marked-up manuscript, and updated page and line numbers are provided below.

## Response to Referee 1

We thank the referee for recognising that the revised manuscript improved the content and framing, and we understand the difficulty created by the scale of the previous revision. In the present response we focus on the specific outstanding points listed in the report rather than introducing another broad restructuring. We have simplified the title and abstract, added concise normalised performance metrics, clarified the evidence limits of the existing scaling experiments, and added a code-level verification of target-ID transfer and recorded coalescence history. We have also revised the five minor wording issues identified by the referee.

### Major comments:

1. **Comment:** The title is improved but perhaps slightly irregular, why is TPHT mentioned whereas forward and backward tracking are not? Perhaps simply “implementation and computational evaluation” is sufficient.

**Reply:** We agree. Mentioning TPHT alone in the title gave disproportionate emphasis to one component of a framework that also includes forward and backward tracking. We have therefore simplified the title to:

“A Lagrangian Particle Tracking Framework for the Super-Droplet Method in SCALE-SDM 5.2.6-2.3.1: Implementation and Computational Evaluation.”

TPHT remains fully described and evaluated in the manuscript, but is no longer singled out in the title.

**Change in manuscript:** title, lines 2-3.

2. **Comment:** I appreciate that you have re-written the abstract with my previous comments in mind, however the new version is jargon heavy and therefore difficult

to understand. I think it serves the purpose of summarising the manuscript well, but only after having read the revised manuscript to understand what all the words mean. Consider using simpler language and explaining any new terms, for example: “shared target-selection”, “sampling subsystem”, “forward discovery pass”, “backward reconstruction pass”, “deduplicated handoff”, “target discovery”, “coalescence-event context”, “target-set diagnostics” and “predecessor-chain reconstruction”.

**Reply:** We agree that the previous abstract introduced too many implementation-specific terms before they were explained. We have revised it so that each workflow is named and immediately described in plain language. Necessary technical terms are retained after explanation, whereas unnecessary labels and repeated shorthand have been removed.

**Change in manuscript:** Abstract, lines 15-24.

3. **Comment:** You may consider moving the analyses into the main text since they are heavily referenced in Section 3.2.3.

**Reply:** We considered moving the complete benchmark and sensitivity analyses into Sect. 3.2.3. However, the main text already reports the principal benchmark and sensitivity results, whereas Appendices A–C provide the complete figures, metric definitions, and evidence limits. Moving the full analyses would substantially lengthen the main case-study section and weaken its focus. We have therefore retained the present division between the main text and appendices and clarified the relevant cross-references. The additional normalised metrics and scaling limitations are addressed separately in our responses to the following two comments. No analyses were relocated; the related additions are described in

Responses 4 and 5 below.

4. **Comment:** It would also be more informative for some metrics in Appendix A to be included which are normalised per superdroplet or per simulated-time-step (e.g. for the wall-clock time and output size).

**Reply:** We agree and have added a compact table of normalised metrics to Appendix A. The new table reports wall-clock seconds per simulated second and per model step, corrected core-hours per simulated minute, selected-output bytes per written target record and per dump, coalescence-log bytes per recorded event, and peak RAM per MPI rank. We define every denominator explicitly and distinguish all-time written records from unique target identities.

**Change in manuscript:** Appendix A and Table A1, lines 961-974.

5. **Comment:** With regards to the “superdroplet scaling” in Appendix C it may be beneficial to measure the scaling as the fraction of tracked superdroplets increases (i.e. from 5% to 10% to 20% etc.) rather than the total number of superdroplets in the domain, or perhaps it can be explained why the scaling would behave similarly in both cases.

**Reply:** We found 5 %, 10 %, and 20 % cases in the archived suite, but these belong to the 600 s quasi-two-dimensional sampling-verification ensemble rather than to a matched three-dimensional performance matrix. Within each sampling mode, the number of written selected records increases approximately in proportion to the retained fraction. However, changes in wall-clock time and peak RAM are not resolved over the 5–20 % range because model runtime dominates and the remaining variation is comparable to the seed-to-seed spread. We therefore do not present these cases as a three-dimensional fraction-only overhead test.

We have also clarified the scope of Appendix C. Its SDNC experiment keeps the tracking fraction fixed at 5 % while increasing the physical super-droplet population. It therefore measures end-to-end scaling as physical-model work, tracking load, and selected-output load increase together; it does not isolate the marginal cost of changing only the tracking fraction. We added both limitations to the manuscript rather than fitting an unsupported fraction-only wall-clock or RAM slope.

**Changes in manuscript:** Appendix B and Appendix C, lines 1030-1034 and 1068-1071.

**Minor comment:**

1. **Comment:** L22: “memory” and “storage” could be the same thing, be clarify (perhaps RAM and non-volatile memory are meant?). Also “output-frequency sensitivity, and sampling behaviour” is vague, please be more specific.

**Reply:** We agree and have replaced the ambiguous terms. The revised abstract now distinguishes peak random-access memory from disk-output volume. We also replaced “output-frequency sensitivity” with “the effect of output interval on file volume and mean write time”, and “sampling behaviour” with “sampling error and seed-to-seed variability”.

**Changes in manuscript:** Abstract, lines 22-23.

2. **Comment:** L60: I am not sure what is meant by “general-purpose” because it’s now clear this tracking package is really only applicable to SCALE-SDM ... Please clarify.

**Reply:** We agree. We removed “general-purpose” from the description of the tracking methodology and now describe the work as a documented tracking implementation for SCALE-SDM. The Discussion retains a narrower statement that some design concepts may be transferable to other Lagrangian particle models if those models provide persistent identifiers, parallel ownership, predecessor links, and model-specific event definitions.

**Changes in manuscript:** Introduction, line 63.

3. **Comment:** L63-74: This paragraph is repetitive with what is written in L39-44, consider restructuring to combine both paragraphs into one.

**Reply:** We agree that the two passages repeated the same literature-gap statement, although they serve different purposes. We therefore retained both contexts but removed the repeated claim. The earlier passage now introduces the implementation requirements of particle tracking in a massively parallel SDM, while the later passage retains the Matsushima et al. (2023) example and states the specific differences from the present work more concisely.

**Change in manuscript:** Introduction, lines 45-47 and 65-71.

4. **Comment:** L87: The sentence: “The framework also includes coalescence-event output and post-processing procedures for [...]” is unclear because it combines two rather separate things, please reconsider what the intended message of this sentence is; “post-processing procedures” is also very vague.

**Reply:** We agree and separated the two functions. The revised text now states independently that the model records event-level collision–coalescence

information and that separate post-processing tools use saved identifiers to reconstruct selected histories from rank- and time-partitioned output files.

**Change in manuscript:** Introduction, lines 81-86.

5. **Comment:** L97: please clarify what “the shared sampling and target-selection subsystem” are when first mentioned, or what they relate to (forward/backward tracking?).

**Reply:** We clarified the term at its first occurrence. It is now defined as the common procedure used by both forward and backward tracking to select target super-droplets through random or stratified sampling and user-defined physical filters.

**Change in manuscript:** Sects. 2.2.1 and 2.4, lines 128 and 290-293.

## Response to Referee 2

We thank the referee for the positive assessment and for identifying the important distinction between target-ID transfer and verification of the corresponding particle histories. We agree that the previous manuscript demonstrated only target-ID transfer. We therefore added a code-level comparison of the recorded coalescence history and a separate discussion of the conditions required for simulation-level reproducibility.

### Major comments:

1. **Comment:** I note that the authors explicitly state that the two passes start from the same physical and random-number state. However, the current consistency check verifies the transfer and recovery of target IDs, but does not demonstrate that the corresponding physical histories are reproduced between the two passes. Although I recognize the authors' efforts to reproduce the identical random state, this validation is not properly discussed in the manuscript.

I would therefore ask the authors to confirm that the two passes remain physically consistent. For example, for a subset of targets, the authors could compare the timing of the first  $r \geq 15\mu\text{m}$  crossing, the radius evolution preceding that crossing, and/or the timing of the coalescence event associated with the target criterion. Such a comparison would provide a direct physical-history validation of the TPHT reconstruction in addition to the existing target-ID check.

**Reply:** We agree that recovery of the same target-ID set verifies target transfer but does not by itself establish reproduction of the corresponding particle histories. We followed the referee's suggested comparison of coalescence-event timing, using the complete retained event logs rather than a small target subset.

A common practical check of simulation-level reproducibility would be to compare

Eulerian output from the two integrations. However, the forward-discovery run did not save Eulerian history or selected-super-droplet state output because these outputs were disabled to reduce storage. Such a comparison therefore could not be made from the archived files. We instead performed a code-level verification specific to TPHT: whether the target-ID list was transferred completely and whether the second integration recorded the same physical coalescence history as the first.

The executable used by both integrations had the same SHA-256 hash, and all 256 corresponding boundary, initial-state, physical random-number-state, and reference-state file pairs were byte-identical. The target-list transfer recovered all 3,554,256 unique identities without missing or additional targets. We then streamed and compared all 26,284 corresponding coalescence-event file pairs, containing 2,919,183 event rows in each integration. Every row agreed bitwise in event time, both pre-update radii, both multiplicities, and the represented collision count (“num\_col”), with a maximum difference of zero for every compared field. The exact agreement of these event records provides evidence that the two integrations reproduced the same recorded coalescence history under the tested configuration. A concise statement of this evidence has also been added to the abstract, while its scope and limitations are explained in Sects. 3.2.4 and 4.2.

The 2,919,183 rows refer to the complete raw coalescence-event logs. This count differs from the 2,919,104 records used in Figure. 3b, which form a derived target-linked subset after applying output-window and backward-chain-linkage requirements. The archived forward-discovery run also contains no selected-super-droplet state histories or first-detection time, radius, and height. Therefore, the first  $r \geq 15 \mu\text{m}$  threshold-crossing time and the preceding radius evolution could not be compared from the retained files. We now state explicitly that the event-log

comparison is a code-level verification of the recorded coalescence history, rather than a complete comparison of all model states.

**Changes in manuscript:** Abstract, Sect. 2.5, new Sect. 3.2.4 and Table 3, Sect. 4.2 and Sect. 5, lines 27-31, 378-380, 606-647 and 893-898.

2. **Comment:** It would also be useful to briefly discuss the numerical reproducibility requirements and limitations of TPHT. Even when the same initial state and random-number state are used, finite-precision floating-point arithmetic, differences in operation ordering, parallel execution, compiler or hardware, or changes in the sequence of random-number calls could in principle lead to divergence between stochastic realizations. Clarifying which aspects must remain identical between the two passes, and to what extent TPHT is expected to remain reproducible across different computational configurations, would help readers understand the practicality of the method.

**Reply:** We agree and have added a separate discussion of the conditions required for the two integrations to follow the same stochastic realisation. These simulation-level requirements are distinct from the target-ID and coalescence-history checks described above. The two integrations must use the same source and executable, compiler and floating-point settings, numerical libraries, initial physical and physical random-number states, MPI rank count and decomposition, effective thread count, floating-point operation ordering, and physical random-number call order. The audited warm-cloud implementation uses a separate random-number generator for target selection, and the identifier, output, and boundary-exchange paths do not feed tracking identifiers into collision-pair construction or reorder the physical super-droplet arrays.

We limit the bitwise statement to the exact computational configuration used in this study. We do not claim bitwise reproducibility across arbitrary compilers, hardware, MPI layouts, effective thread counts, code versions, or changes in floating-point operation ordering or random-number call sequence. Because Eulerian and selected-super-droplet state output was disabled in the discovery run, we also do not claim that every later model-state variable was directly compared.

**Changes in manuscript:** Sect. 2.5 and Sect. 4.2, lines 398-404 and 843-853.