Referee's Comment on "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"

Title 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
Author(s) Chongzhi Yin, Shin-ichiro Shima, Chunsong Lu
MS No. egusphere-2025-6221
MS type Development and technical paper

General Comments

This paper details the algorithms which have been implemented in SCALE-SDM to backward/forward track superdroplets. Most notably, in my opinion, the paper explains how the step for looking-up superdroplets in the backwards tracking algorithm is only $\mathcal{O}(1)$ complexity. The tracking algorithms are demonstrated complementarily using DYCOMS-II RF02 large eddy simulations and whilst the results of the test-case are not ground-breaking or especially detailed, they nevertheless serve the purpose well of demonstrating the usefulness of both algorithms. The manuscript is well-written and on the whole very clear, even if at times a little bit "AI-ish" and bold.

I commend the authors for carefully documenting the development of their code and highlighting its particular innovations in this manuscript. Such efforts are rarely undertaken in the cloud microphysics community (I am fairly sure other super-droplet modellers have their own tracking algorithms in some shape or form, yet this is the first manuscript to the best of my knowledge which details them); the knowledge shared is of great benefit to the community and leaves many potential avenues for further development, as outlined in the manuscript's discussion. Rather than anything depreciative, the many points I have made in this review indicate the high quality of the manuscript and that I could understand and engage with it.

Notwithstanding that the manuscript is worthy for publication, I see three points on which I believe it could be majorly improved:

- the paper lacks a sufficient computational evaluation of the algorithms. A thorough evaluation of the memory footprint of both algorithms, the efficiency with which they use computational resources, and their scaling with increasing number of superdroplets would greatly enhance the claims the paper makes along the lines of computational efficiency. Discussing the computational efficiency in the context of the same simulations without tracking would also be extremely beneficial here. Without a more thorough analysis of the computational expense of both algorithms the paper still serves a valid purpose, to comprehensively document the algorithms for the community, but then the expectations and key statements made by the paper should be re-aligned to match this (see specific comments for particular points in the manuscript where this applies, mostly on statements claiming the algorithms are computationally efficient.).

- It is unclear what the innovations are in the forward-tracking algorithm. I understand the backward-tracking method is particularly fast because of its $\mathcal{O}(1)$ lookup of superdroplets. In contrast, to me the forward algorithm looks like a fairly standard approach to selective data output from Lagrangian simulations. Is the method for sampling of superdroplets particularly special? If there is a particular innovation in the forward algorithm it would greatly benefit from being more clearly highlighted. Else, if this algorithm is indeed already a rather standard approach, the paper should make clear what exactly is the goal of documenting it here, e.g. to highlight the particular details of the SCALE-SDM version.

- Are the algorithms specific to superdroplets as opposed to any sort of Lagrangian particles/parcels? I suggest rephrasing mentions of superdroplet tracking to Lagrangian particle tracking or simply particle tracking. Even in the title for example "for the Super-Droplet Method" could be removed. If there is something about these algorithms which makes them specific to superdroplets, aside from their implementation in SCALE-SDM, please state this more explicitly. If not, rephrasing superdroplets to Langrangian particles would greatly increases the scope of this manuscript, widening it's appeal to Lagrangian modellers outside superdroplets and clouds microphysics too.

Specific Comments

There are several sentences in the introduction which appear to be vague and/or misleading:

- **P2 L45-46:** consider rephrasing the sentence ending "unique challenges and opportunities". This phrase is vague and not addressed in the paragraph that follows.
- **P3 L67-69:** Rephrase this sentence, at least remove "therefore", but also reconsider the term "computationally efficient" (as per general comment above and because it is somewhat contradictory with the 3-D tracking simulation not being possible due to computational expense).
- **P3 L80-83:** Reconsider description of the tracking algorithms here as "robust", "computationally efficient" and "designed for broad scientific inquiry"; it is not clear how the algorithms presented are robust or broad, please either explain what is meant by these terms somewhere in the manuscript or rephrase this. The term computationally efficient is misleading (as above), and this is also the case in **P19 L601**.
- **P3 L88:** I have trouble with the phrase "statistically robust" to describe the forward tracking algorithm. On **P12 L360-361** a 5% sampling is used and it is not immediately obvious, nor shown statistically that this corresponds to a robust sampling of the population. On **P19 L572** the term "statistically representative" is likewise problematic.
- **P3 L91-93:** "Other researchers . . ." please be much more specific, consider using one or two example(s) which may be more effective at conveying what is meant than such a general sentence. Similar point for the sentence **P18 L592-595**.
- **P4 L98:** I find "presenting key scientific insights" misleading or overstated, consider rephrasing or being more specific.

Both algorithms would greatly benefit for some clarification and discussion of the files they create:

- from what I understand the backward algorithm creates a very large number of separate small files. Is this correct? And what is the structure of these files and was there any consideration on how big/small they should be to optimise reading and writing them? Perhaps memory can be saved by reconsidering how these files are written/organised too.
- If the pre_sdid and pre_dmid are also written to file, please add them to **Fig. 1**.
- It may be worth already mentioning in the backward algorithm's description that it is expected to have a very large memory footprint given how much data it outputs, as is then proven later in the manuscript for the DYCOMS II test-case.
- In the forward algorithm (**P9 L260**), what are the array dimensions? If they are something like (Time, Domain ID, Superdroplet ID) they would be very memory intensive and could be better optimised, perhaps this too could be pointed out.
- On **P18 L570** you might consider data compression, potentially with AI, as a useful strategy to pursue in the future in order to reduce the algorithms' memory costs.

Some additional clarification on a couple of details of the backward algorithm would also be helpful:

- the lookup of superdroplets is $\mathcal{O}(1)$ and so I believe this means it is no longer the most critical scaling of the algorithm. Does the complete algorithm hence scale linearly with superdroplets? Or perhaps in some way with the batch size of the post-processing jobs?
- On **P7 L212-213:** the statement after " to further optimize" is not an optimisation, it is a description of ordinary cache-loading in computers. Please reconsider this statement.
- **P13 L401:** Can you estimate the memory footprint of the coalescence logs that would've been produced if coalescence had been activated? It seems important information for a user who would use the algorithm with coalescence enabled. Or perhaps this could be clarified by a more detailed discussion of the files the algorithms output as commented on above.

Likewise I think some clarification is necessary of the following points related to the forward algorithm:

- a call to ADD_RANDOM_PERTURBATION_ALL occurs which doesn't occur in the backward algorithm. Why is this? Consider stating it is not essential to the algorithm if it is not, or explain why is it necessary for the forward (but not backward) algorithm.
- Sorting is mentioned in post-processing, however non-specific sorting algorithms are often extremely costly and so the choice and cost of the algorithm should be elaborated on if the computational performance of the algorithm is further presented (as commented on above).

Finally, please elaborate on the **author contributions**, CL is not even mentioned.

Technical Corrections

- **P1 L16:** delete "model" from "Large-Eddy Simulation model"
- **P2 L49-50:** rephrase "compelling scientific outcomes", this is vague and obscure.
- **P3 L74:** quote unnecessary, can be removed, also correctly cite ''their Figure 6".
- **P3 L85:** consider rephrasing "The importance of this contribution is significant".
- **P4 116:** "the SDM" include SDM version and omit "the" (as done with SCALE in the paragraph before). "a" and "the" before acronyms are generally omitted, consider applying elsewhere in manuscript too.
- **P5 L135:** fix formatting which started new paragraph unnecessarily.
- **P5 L144:** please define a "sequence number". In the same paragraph it would also be helpful to the reader to explicitly state somewhere that "pre_sdid" stands for pre_superdroplet_identity and "pre_dmid" stands for pre_domain_identity.
- **P5 L145:** please use a more common word for "antecedent".
- **P5 L154:** Is the if_coal flag always set to 1? Also in **Fig. 1** if_coal = 1 makes it seem like it is always 1?
- **P6 L173-174:** omit the redefinition of multiplicity.
- **P6 L182:** "referring the . . . " sentence grammatically incorrect; missing "to"?
- **P7 L195-197** repetitive, consider omitting.
- **P7 L207:** memory usage and I/O bottlenecks are not typically "balanced", please rephrase (e.g. with "reduce"?).
- **P7 L209:** typo, replace one "pre_dmid" with "pre_sdid".
- **P8 L 230:** does the INVALID_ID consume memory? If so, is it transferred by MPI communication? Both these things could be potential sources of (minor) optimisation of the forward algorithm.
- **P9 L275:** Please elaborate on how utilising a database would save memory, or rephrase.
- **P9 L287-289:** This definition of eddy-hoping is not quite correct, please consider something more along the lines of Grabowski and Abade 2017 "droplets arriving at a given location within a turbulent cloud follow different trajectories and thus experience different growth histories and that this leads to a significant spectral broadening" (and/or the definitions in the other two papers cited here).
- **P9 L289:** "Furthermore this method", state "backwards method" for clarity.
- **P11 L337:** fix "(Nishizawa et al., 2015)" citation formatting.
- **P11 L353:** please explain how momentum exchange between super-droplets and the fluid is taken into account, or provide a reference.
- **P12 L360-361:** use consistent formatting for 400,000 and 7,950,000, preferably standard form. Consider also stating this is roughly 5% of the population.
- **P12 L373-374:** consider putting parentheses around "The current version of the model . . . y-direction".
- **P12 L381: Table 1** states coalescence was active but here it is disabled, please correct this inconsistency.

- **P13 L398:** Why are there more output intervals for the backwards than forwards tracking?
- **P13 L404:** What is a "standard workstation"? Please define or be specific.
- **P13 L406:** I think "drastically" in this context is unproven, provide evidence/explanation or simply say "could reduce the runtime."
- **P14 L439:** "trends" is not the right word here, consider "findings" or "results" instead.
- **P15 L457:** "similar" is slightly misleading, consider "within a factor of 10"?
- **P16 L484:** unnecessary to say "on the Horizontal distance-Height plane", it is already in figure caption and axes labels. Likewise you can remove the sentence thereafter. Also in figure captions, height and horizontal do not need capitalisation.
- **P16 L504:** "strong [upward] vertical velocity" or is it downwards too?
- **P16 L507:** "The selected super-droplets at this time (see Figure 7)", is the reference to figure 7 correct?
- **P17 L531:** I'm not sure "novel" is an accurate description, consider simply "provided insights".
- **P18 L545:** rephrase "significant pathway", "significant" in this context is vague.
- **P18 L548:** "impossible to uncover" is too strong. Many previous works, including ones cited in this manuscript, did not use this backward tracking algorithm and nevertheless uncovered the lifecycle of individual particles.
- **Algorithm 1 and Algorithm 2:** both algorithms mention detailed implementations in the Appendix, please provide these or remove this statement. Also fix the formatting with/without colons in the algorithm labels.
- **Fig. 2:** Are Q1, Q3 and the mean almost identical in (a) and (b)? Plotting these on the same plot (a small subfigure (c)?) would aid comparison.
- **Fig. 4 and Fig. 5:** consider plotting figure 5 on top of figure 4, or at least with the same axes limits and scales. As it stands, I am not sure that Fig. 4 adds significant value to the manuscript and so I think it could be removed entirely.
- **Fig. 7 and Fig. 8:**: Please label with minutes instead of seconds to be consistent with the text and other figures.