Review of the GMD-2019-343 manuscript

Hereby I provide my comments to the manuscript entitled "Collection/Aggregation in a Lagrangian cloud microphysical model: Insights from column model applications using LCM12 (v0.9)"

Overall, I consider the submitted text a valuable contribution to the literature on particle-based cloud modelling as it provides detailed formulation of test cases that are essential in the development of new implementations of relevant algorithms. It thus clearly matches the journal scope.

Below, I list my major, minor and technical comments to the manuscript. I include also a comment to the enclosed software.

Major points

Unacknowledged performance trade-offs

Some performance trade-offs pertaining to the choice among linear and quadratic sampling are detailed. It is however not pointed out that linear sampling precludes parallelisation of the collision computations (within a gridbox/column) due to introduced data dependency (page 11, line 22-23). This is particularly worth underlining, as the availability of shared-memory parallelisation with multi-core CPUs or GPUs allows for significant speed up (i.e., almost by the factor equal to the number of threads). In fact, all but the pair-shuffling and random number generation steps in the AON coalescence algorithm with linear sampling are embarrassingly parallel. Given the above, I find it at least misleading to say, without mentioning the precluded parallelisability, that:

- "simulations with linear sampling ... converges slower ... compared to quadratic" (p1/l14-15)
- "benefit of the reduced computational cost may be outweighed by the stronger requirements on Δt" (p24/l2)
- "restrictions on the timestep might cancel out the computational benefit gained by the reduced number of SIP combinations" (p39/l25-26)

Similarly, I doubt the statement on page 12, lines 2-3 (on performance superiority over integer-preserving implementation) holds true in parallel context, where random numbers can be generated concurrently in large batches.

Finally, it is worth commenting on the parallelisability consequences of the requirement to perform collisions column-wise in the WM2D scheme.

Subrid-dynamics and WM2D

While it is acknowledged on page 14 (lines 15-17) that the WM2D scheme is somewhat incompatible with "sophisticated kernels", I highly recommend to extend the discussion also to the aspects of subgrid-scale dynamics representation in particle-based models – e.g., referring to the already cited work of Hoffmann et al. (2019). In short, in my understanding, the "information content" of SIP positions in particle-in-cell-type models is somewhat overestimated here. In particular, the prevalent Large-Eddy-Simulation context should be addressed. Candidate location: page 14 (lines 11-14)?

Paper length

The article length is, in my opinion, impeding appreciation of its content. I include some detailed suggestions on what could be omitted from the text in Technical/editorial remarks below. Besides that, I consider it a malpractice to introduce an almost-page-long quote from an earlier study of the authors. I see also little benefit in repeating Figure. 3 here – please just refer to the relevant parts of the 2017 GMD paper which is readily available for all readers.

I also suggest adding a table of contents (as done recently in GMD in Shima et al. [3]).

Minor points

The title

First, why not to avoid a "slash" in the title, and use "Collisional growth" instead of "Collection/Aggregation". Second, I oppose to calling presented simulations "applications", suggest "simulations"?

Finally, I generally suggest to label the discussed microphysics modelling methods as "probabilistic particle-based" rather than "Lagrangian". First, "Lagrangian" is a much more broader term (consider e.g. the Lagrangian cloud models described in [2, 1]), and thus potentially misleading for readers from outside our niche. Second, the discussed model is not fully Lagrangian as it relies on Eulerian dynamical core.

I am aware that the present title is a reference to the U2017 paper, but perhaps the above arguments outweigh it?

On a related note, there is not a single mention of the "Monte-Carlo" keyword in the paper, please do cater to a wider community and use such keywords to give a good context for the readers.

The limiter

The ad-hoc definition of the "limiter" (p15/l6) calls at least for a reference to the min() in the SDM paper's $\tilde{\gamma}_\alpha := \min(\gamma_\alpha, [\xi_{j\alpha}, \xi_{k\alpha}])$ expression (Shima et al. 2009, step (5) in the left column), if not for reformulating the "limiter" in a more robust manner.

Integer vs. real-valued weighting factors

I do not find enough grounds in the text for the statements on the superiority of real-valued vs. integer-valued weighting factors. Besides the above commented issue of parallel random number generation/multiple collisions, the statements on page 11, lines 30-31 seem to overlook the concept of spectrum estimation, see e.g. the third paragraph in section 5.1.4 in (Shima et al., 2009) (also, worth mentioning when discussing eq. 14).

CFL condition for sedimentation

I would argue that we should assume the CFL condition for particle sedimentation as well – while it does not cause the numerics to fail as in Eulerian component, it is intuitively not fulfilling the assumptions (as acknowledged on p4/l12). Relevant statements: p8/l11, p12/l26, p21/l15, p34/l10.

Courant number values in section 3.1

Upwind and MPDATA convergence is dependent on the Courant number – please indicate which was used in section 3.1. Perhaps worth checking the behaviour for a set of Courant numbers.

Correlations

Please be more specific as to the mathematical meaning of "correlations" mentioned 12 times in the context of collisions but never defined.

Classical/regular nomenclature

References to "classical/regular" "implementations/approaches/cases/versions/AON" (p8/l15, p8/l16, p8/l22, p8/l24, p9/l4, p11/l12, p12/l8, p12/l19, p12/l29, p12/l31, p14/l7, p14/l8, p14/l30, p15/l4, p17/l2, p19/l1, p25/l5, p25/l21, p29/l1, p30/l2, p37/l12) are not understandable, especially given that the authors introduce their own nomenclature for numerous notions named differently in literature. I am aware that there are some definitions of "regular" in the text, but it is an over 40-page long paper. Please come up with more precise and less subjective statements.

Technical/editorial remarks

Figures

I urge the authors to replace raster images in figures 1 and 4-23 by its vector-format equivalents (i.e., `plt.savefig(format='png', ...)` $\rightsquigarrow$ `plt.savefig(format='pdf', ...)`).
I suggest using cm^{-3} as the unit for λ_0 on the plots.

Text

p1/l2 the "high number" is equally (or even more) applicable to bin models, please rephrase and indicate with respect to which benchmark the value is high

p1/l4 is the word "explicit" needed (suggest avoiding if the opposite "implicit" is not clear)

p1/l8 ditto

p1/l12 "accuracy" $\rightsquigarrow$ "resolution"

p2/l28 please underline that it is you who introduce the AON term

p2/l22 why not in chronological order? (see also background works listed in Shima et al. [3] and dating back to 2004)

p2/l26 "abbreviated as ... in the following" $\rightsquigarrow$ "hereinafter abbreviated ..."?

p3/Table 1 please remove non-English caption, and consider removing the table – given the multitude of symbols used in the text, it seems anecdotal to list 8 abbreviations in a table (moreover, the following are not listed: DNC, MPDATA, CFL, WM2D, WM3D, MC, BIN, US1, noSedi, LCM0D, LCM1D, LS, ...)

p3/l3 suggest removing/rephrasing "relatively young modelling approach" – particle-in-cell method is 50-year old; same for Monte-Carlo for collision

p3/l4 "aws" $\rightsquigarrow$ "was"

p3/l12 "coalescence, aggregation, or accretion" – mention "riming", "self-collection", "wash-out" keywords as well?

p3/l13-14 actually, ice crystals are mentioned only one after this statement – suggest removing

p3/l22 "(sometimes pedantically)" sounds negative ... suggest not being pedantic (see below) and removing the statement

p4/l3 suggest using subscripts for L_z and n_z

p4/l6 no need to make it a separate numbered equation?

p4/l8 no need to define volume of the sphere - just mention in the text

p4/l12 $K(m_i, m_j)$ and $K(r_i, r_j)$ would likely be better named differently

p4/l15 "radius-dependent" $\rightsquigarrow$ "sie-dependent"

p4/l17 "latter" $\rightsquigarrow$ "last" (there are three assumptions listed before)

p4/l18 there is a bogus character before 500 μm - garbage displayed in the pdf viewer I'm using

p5/l1 "collection" $\rightsquigarrow$ "collisional growth"?

p5/l9 "latter" $\rightsquigarrow$ "last" (there are three assumptions listed before)

p5/l12 no need to define factorial

p5/l13 skip reference to Berry 1967, surely "mass density function with respect to the logarithm of radius" is enough

p5/l22 first sentence is a repetition from the Introduction

p5/l24-25 suggest rephrasing around "terms low and high"

p6/l14 random interval should be [...) and not [...] (as per numpy.random docs)

p6/l14 "some threshold" – please be specific (also, worth mentioning the alternative formulation with integers)

p7/l3 "around" – please be specific

p7/l6 random number interval: [...)

p7/l6-7 unneeded sentence (Furthermore...)?

p7/l15 rephrase "to solve the Smoluchowski."

p7/l18 use partial derivatives

p7/l21 constant-in-altitude air density implied – please mention (same concerns the assumption that w_{sed} is constant

p7/l24 the two Smolarkiewicz papers list several flavours of MPDATA, which is used?

p7/l26 "some value" $\rightsquigarrow$ "the value"?

p7/l30 why 0.5?

p8/l10 "Unlike to" $\rightsquigarrow$ "Unlike in"

p9/Fig 2 caption: "Wellmixed" $\rightsquigarrow$ "well-mixed" (twice)

p10/l4 is there any added-value in including the time loop in the pseudocode (same concerns listing on p13)

p10/l5 indicate that only for quadratic sampling

p10/Alg. 1 caption random number interval: [...)

p11/l11 random number interval: [...)

p12/l29-31 use a proper big-oh (e.g., $\mathcal{O}(n)$ with `$\mathcal{O}(n)$`)

p13/Alg. 2 caption random number interval: [...)

p14/l2 use a proper big-oh; nz inside the oh?

p14/l18-21 needed?

p15/l19 comment that constant air density implied

p16/l12 remove "in the column model source code"

p16/l16 omit "Validation exercises" from the section title (entirety of the paper is a validation exercise)

p17/Fig. 4 "at the indicated points in time" – cannot see any indicated points in time

p18/Fig. 5 ditto

p18/Fig. 5 "use different y-axis" – cannot see different y-axis

p18/l2 first mention of "lucky droplets" calls for a reference

p19/Fig 6 BoxModelEmul not mentioned before, used only in figure captions

p19/l6 "surprisingly well" – please be more specific

p19/l7-8 "usually more easily reached" – please be more specific

p19/9-10 "Even though..." – suggests this material can be skipped

p19/l12-14 sounds like a sentence for Introduction or Conclusions

p19/Fig. 7 no units for dV (BTW, shouldn't it be ΔV?)

p20/l3 "according" $\rightsquigarrow$ "relevant"?

p20/l8 avoid "We believe"

p16/l31 "agreement ... is good" – please be specific, provide quantitative measure

p21/l24 "In a technical experiment" – suggest rephrasing

p21/l30-32 "Nevertheless..." – suggest skipping the sentence

p22/l24 first mention of Long kernel without reference or comment

p22/l30 rephrase around "now f or"

p23/Fig. 8 no units for "dt" (BTW, shouldn't be Δt?)

p23/l3 "occur too often" – please be more specific

p24/l3-4 "has to be solved" – subcycling seems to me as a preferable option than changing timestep of the whole simulation ... suggest skipping/rephrasing the sentence

p24/l4-5 please elaborate how/why inclusion of more SIP attributes would change the influence of LinSamp vs. QuadSamp choice?

p25/Fig. 10 add ", respectively" at the end of caption

p25/l13 "Wang" $\rightsquigarrow$ "Wang et al."?

p25/l8 good place to mention performance trade-offs of the WM2D

p25/l21 Bott/Bott's algorithm/model – please be consistent

p27/l5 "Bott's results are reliable" – rephrase

p28/l7&8 suggest renaming the section to "Algorithm profiling"

p29/Tab. 1 use the same exponential notation as elsewhere (i.e., $A \cdot 10^B$ instead of AeB)

p30/l6 "find their" – rephrase

p30/l23 "For small SIPs j" – what is a small SIP?

p31/l24 unit of "influx" should include 1/time

p32/Fig. 15 avoid using two acronyms for the same thing: LS, LinSamp

p33/Fig. 16 harmonise case for acronyms: Bin, BIN

p33/l2 "collector SIP" – please elaborate

p34/l4 "sounds like a banal ..." sounds too colloquial

p34/l14 what is the former effect?

p35/l4 "superiority ... in BIN" – rephrase

p35/l12 "improvement of this" – rephrase

p37/l14 same remark regarding riming, etc as for p3/l12

p37/l11-12 please clarify if this statement concerns just this section

p39/l11 "To bridge the gap" – puzzling, the gap was not mentioned earlier

References

Please harmonise the reference format:

- avoid double URLs (DOI URLs are unambiguous and enough): p41/l3-4, p41/l6, p41/l12-13, p41/l19-20, p41/l24-25, p41/l27, p41/l30-31, p42/l2, p42/l7-8, p42/l12-13, p42/l26, p43/l4-5, p43/l26-27
- add DOIs:
  - Kessler 1969: 10.1007/978-1-935704-36-2_1
  - Khairoutdinov and Kogan: 10.1175/1520-0493(2000)128%3C0229:ANCPPI%3E2.0.CO;2
  - Matsumoto and Nishimura 1998: 10.1145/272991.272995
  - Naumann and Seifert 2015: 10.1002/2015MS000456
  - Seifert and Beheng 2001: 10.1016/S0169-8095(01)00126-0
  - Shima et al. 2009: 10.1002/qj.441
  - Simmel et al. 2002: 10.1016/S0169-8095(01)00131-4
  - Smolarkiewicz 2006: 10.1002/fld.1071
  - Smolarkiewicz 1984: 10.1016/0021-9991(84)90121-9
  - Sölch and Kärcher 2010: 10.1002/qj.689
  - Wang et al. 2006: 10.1063/1.1928647
  - Wang et al. 2007: 10.1016/j.jcp.2007.03.029
- remove months: p41/l18, p41/l21, p41/l24, p41/l26, p41/l30, p42/l6, p42/l12, p42/l25, p43/l14, p43/l37, p44/l2
- remove ISSNs: p41/l24, p42/l26, p43/l14, p44/l2

Please correct:

- volume number in Grabowski et al. 2019
- page range for the work of Smoluchowski, should be 557-571 (see `https://jbc.bj.uj.edu.pl/dlibra/doccontent?id=387533`)

Enclosed software

I consider it awkward and confusing to use the GCC preprocessor with Python source code. Same concerns using csh and sed to generate preprocessor-directive-filled .py files with hardcoded system-dependent paths. Altogether, the multi-platform and work-out-of-the-box advantages of Python were lost.

Note that in the `CompSim.gcc.py` file you are using two independent random-number-generation infrastructures available in Python, and I doubt setting the seed via `random.seed()` affects values returned via `np.random.random()`.

Please avoid non-English comments in the code and marking changes with comments – it is the role of version control to track changes.

Please indicate in the code availability section:

- the code availability (or lack thereof), version and license for "Bott" and "Wang" models
- the license the LCM1D is released on
- the supported environments and dependencies of the implementation (python, numpy, gcc, csh, sed, ...)

Thank you for a useful contribution to the field!
Hope the above comments help,
Sylwester

References

[1] D.G. Dritschel et al. "The moist parcel-in-cell method for modelling moist convection". In: *Quarterly Journal of the Royal Meteorological Society* 144 (2018). DOI: 10.1002/qj.3319.

[2] S.G. Lasher-trapp, W.A. Cooper, and Alan M. Blyth. "Broadening of droplet size distributions from entrainment and mixing in a cumulus cloud". In: *Quarterly Journal of the Royal Meteorological Society* 131 (2005). DOI: 10.1256/qj.03.199.

[3] S. Shima et al. "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". In: *Geosci. Model Devi. Discuss.* (2019). DOI: 10.5194/gmd-2019-294.