Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., author comment AC2
https://doi.org/10.5194/gmd-2021-125-AC2, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC2

Camilo Fernando Rodríguez Genó and Léster Alfonso

Author comment on "Parameterization of the collision–coalescence process using series of basis functions: COLNETv1.0.0 model development using a machine learning approach" by Camilo Fernando Rodríguez Genó and Léster Alfonso, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-125-AC2, 2021

The authors thank Anonymous Referee # 2 for her/his helpful comments that improved the quality of the manuscript.

Overview of the revised manuscript:
A general revision of the draft article was performed, including changes in its content, resulting in a slightly longer, more comprehensible draft. A summary of the main changes is included as follows:

- The Introduction has been restructured, to provide more clarity about the state of the art and a better understanding of the main ideas of the article.
- A complete grammar review of the article has been done, resulting in a paper more friendly to the reader
- Figure 4 has been discarded, as it duplicated information from Table 1.
- An extra parameterization has been added, in order to compare its results with those of the Machine Learning parameterization.
- Several figures have been modified to reflect the inclusion of the additional parameterized model.
- The conclusions are now supported with the analysis of the comparisons of three models, instead of two.

Comments from Anonymous Referee # 2 and answers from the authors

General comments

I think that the core idea of the paper, to replace a computation in the simulation of the collision coalescence process with the predictions of a machine learning model, is a valid one. However, I have a few major concerns and questions:
• First of all, the manuscript is in need of some thorough editing for clarity and correctness. There
are plenty of grammar mistakes, typos, and confusing phrasing, such that it is overall not pleasant to read.
• Given that the machine learning application presented here is very straightforward (the training

data cover all possible parameter ranges the model will encounter in the experiment, so all the
model has to do is to learn how to interpolate the training data, there is no generalization needed
beyond what it has already seen), I would have wanted to see a better justification of its utility.
More concretely: How much time and/or memory is saved by the DNN compared to directly
computing the moment tendencies (Eq. 13) using a numerical integration method such as a
trapezoidal rule, and compared to using a lookup table for these integrals (the introduction
mentions that this is a commonly used method)? It would also be interesting to see how these
time savings compare to the total runtime of a typical simulation (since runtime optimization
should aim at the computational bottlenecks).
For example, a lookup table of the size of the dataset used here can fit in a Level 3 cache (if I
understand correctly, 1'000'000 samples, so 1'000'000 x 5 targets were generated in total –
assuming each target is a 64 bit (8 byte) float, we get a total size of about 40 MB), so it might
well be that the lookup table is faster than the DNN predictions (but of course, it requires more
memory, and it only contains moment tendencies for a pre-defined set of input values whereas
the DNN will predict on any given input). Without estimates of the trade-offs (accuracy, speed,
memory demand) involved, it is impossible to see the added value of using a machine learning
model for the task of predicting the moment tendencies.
• I think the study would be stronger if the new parameterization was not just evaluated for a
single experiment, but for several experiments with different initial conditions, maybe even
exploring some of the "edge cases" (e.g., what happens when the number of drops approaches
1, which is the state any collision-coalescence process will converge to?)

Answer:

- Regarding the need of thorough editing, we have performed a major review of the grammar and phrasing, thanks to the helpful comments of both referees, and the reviewed version of the manuscript should have improved in quality.
- Regarding the justification of the utility of the machine learning parameterization, it relates to the more straightforward way of computing the moment tendencies mentioned in the manuscript. Since the numerical solution of eq. (13) is a complex task, particularly the selection and implementation of an efficient numerical method or quadrature for the solution of double integrals, and as the use of lookup tables is a popular but less-than-ideal solution of the problem, the objective of the manuscript is to find an alternate way of computing the rates of the total moments, without sacrificing precision. An exhaustive computational or hardware-focused analysis of the problem falls outside the scope of the presented paper, since the performance of the parametrization depends specifically of the computational platform employed to run the simulation, and the characteristics of the hardware. Besides, as the model is not coded in parallel, it would make no sense to evaluate those characteristics, because it would

not be using the full potential of the computational platform employed, and the distribution of the processors (including caches) and memory flow is in a single way.

- Regarding the realization of new experiments, the authors agree with the referee on performing more test simulations. However, it is not the objective of the paper to show the behavior of the parametrization under several initial conditions, or even under extreme (edge) cases of study, but to introduce the Machine Learning methodology applied to the series of basis functions modelling philosophy, and to eliminate the need to solve complex integrals as part of the formulation of the parameterization. Further testing will be done addressing those and more concerns, including the addition of a condensation module to the parameterization.

Specific comments

L 9: "drop spectrum", not "drop spectra" (it's singular)

Answer: The error has been fixed.

15: "stablish" should probably be "establish"

Answer: The error has been fixed.

L 23: "who used" instead of "whom employed"

Answer: Fixed

L 24: "has shown", not "have shown"

Answer: Fixed

L 27: "For spherical particles such as cloud drops, a transformation of the DSD leads to a self preserving form" – can you briefly explain what this means? Also, it is unclear how this and the following two sentences connect to the previous sentence, which highlights the superiority of
the lognormal distribution in terms of squared-error fit compared to gamma or exponential distributions.

Answer: The order of the sentences in that first paragraph of the Introduction was mixed. To avoid further complications or missunderstandings regarding this section of the Introduction (that was picked up by both referees), that section has been removed from the Introduction.

L 28: Maybe remind the reader of the definition of the Knudsen number and its implications for
the validity of the continuum assumption of fluid mechanics?

Answer: Same as previous answer.

L 25 – 34: I find the purpose of this whole segment unclear and its phrasing confusing. Is the
idea to underline the suitability of the lognormal distribution to the modeling of cloud droplet
size distributions? If so, please make this more explicit and state when a sentence is specifically
about lognormal distributions. E.g.,"The analysis of […] showed that the lognormal distribution
adequately represents the particle distributions" seems to be aimed at strengthening the

case for
the lognormal distribution as an adequate description of DSDs (it needs a citation though), whereas the following sentence ("Further, …") seems to be a general statement about the dependence of the rate of convergence on the initial geometric standard deviation.

Answer: Same as the two previous answers.

L 36: The abbreviation DSD has already been introduced in L 21.

Answer: The second definition of DSD has been deleted.

L 44: "need to calculate a huge amount of equations, which number ranges from several dozens
to hundreds, at each grid point and time step" –> "need to calculate dozens to hundreds of
equations at each grid point and time step"

Answer: Fixed.

L 44: Also mention numerical diffusion as one of the major problems with bin microphysics?
See e.g. [1]

Answer: While it is true that one of the major problems with bin microphysics, and microphysical calculations in general is the numerical diffusion, it is highly dependent of the numerical method used to solve the KCE. For example, the method used (Bott, 1998) is specifically designed to be mass- conservative and to limit the natural diffusiveness of the problem at hand. However, an explanation on this matter is included in the revised version of the manuscript.

L 57: "20 μm and 41 μm being" instead of "being 20 μm and 41 μm" – I won't continue to do
"micro-corrections" of grammar and typos, but the manuscript really needs some thorough editing for clarity and correctness (see my first general comment). Not being a native English speaker myself, I do understand the difficulty of writing in a foreign language, but putting some effort into this will result in a more reader-friendly paper that stands a better chance of getting read and cited by other scientists.

Answer: Fixed.

L 91: This introduction to machine learning seems kind of out of place, especially after the
previous paragraph already talks about deep neural networks.

Answer: The authors agree with the referee, and the paragraphs have been switched to provide more clarity for the reader.

General remark about equations: Please define all variables involved, even if their meaning
seems straightforward – e.g., in Eq. (1), say that r is radius, in Eq. (2), say what N is, etc.

Answer: Fixed.

Neural network architecture: How did you come up with this specific architecture? Did you try
other (e.g., simpler) architectures as well?

Answer: Initially we tried a conventional feed-forward network, very similar to the one used in (Alfonso & Zamora, 2021), which is simpler and the training process is a lot faster. The results with that architecture were good. Taking that as a base, we move forward to try different types of neural network architectures, and we learned about the cascade-forward architecture. We decided to test it and select the one with the best results. Using cascade-forward networks was a time-consuming task, but worth it in the end, as the results improved in accuracy in at least two orders of magnitude using the same number of neurons.

L 171: The commonly used terminology in machine learning is that the training data are the data used to fit the model, the validation data are used for model selection (e.g., when you are testing different neural network architectures, or comparing, say, the neural network with a random forest model, you decide on a final model based on the models' performances on the validation data), and the test set is used for assessment of the generalization error of the final chosen model (see e.g. [2]). Since no model selection is done in this study, what is the called "validation set" should more appropriately be called the test set here.

Answer: The validation set have been renamed test set in the manuscript.

L 214: How were the ranges of the μ and σ parameters (rightmost column of Table 1) for the uniformly random sampling of the distribution parameters that was used to generate the training data determined? Were they "reverse engineered" based on a certain range of LWC values that are thought to be physically reasonable?

Answer: The ranges were determined partially based on data from the CRYSTAL-FACE experiment mentioned in (Alfonso & Zamora, 2021). From that point onwards, we extended the ranges in order to cover a very extensive parameter space complementing the ranges with data from previous simulations using the original parameterization.

L 234: I think it would be interesting to include the collision-coalescence parameterization using the trapezoidal rule to solve Eq. (13) in the results (e.g., in Figure 8) – presumably the main advantage of predicting the moment tendencies using the DNN rather than computing them using the trapezoidal rule is computational efficiency, so it would be nice to know how much faster the DNN is, as well as to see how the mass density spectra obtained using this "trapezoidal parameterized model" compare to those shown in Figure 8 (reference solution and predicted parameterized model). See also my second general comment. Based on the good agreement between the DNN predictions and the validation targets computed using the trapezoidal rule (Figure 7), the resulting mass density spectra will probably look very similar, but I think it would still be interesting for the reader to see that comparison.

Answer: As it is correctly though by the reviewer, the results of the original parameterization and the ML-based model are similar enough that should not be included in the manuscript, to avoid repetition. The main advantage that offers the use of ML is the simplification of the procedures to solve eq. 13, which is very complex to solve numerically, with the exception of using very costly numerical schemes. For instance, the standard quadrature does not apply to eq. 13, and the use of lookup tables is not among the best solutions to the problem.

L 336: I think it's a bit of a stretch to say that the third mode in the evolution of the KCE generated spectra "is reproduced by the parameterization as a wider second mode" – it seems to
me that the parameterization is not able to capture that development.

Answer: The phrasing has been changed to reflect that fact.

Figures
7: The x axis label ("Actual Total Moment Tendencies") of M0 and M1 are missing

Answer: As M0, M1, M4 and M5 share the same x-axis label, it was omitted in M0 and M1 to avoid an overload.

References
Alfonso, L., & Zamora, J. M. (2021). A two-moment machine learning parameterization of the autoconversion process. Atmospheric Research, 249, 105269. https://doi.org/10.1016/j.atmosres.2020.105269
Bott, A. (1998). A flux method for the numerical solution of the stochastic collection equation. Journal of the Atmospheric Sciences, 55(13), 2284–2293. https://doi.org/10.1175/1520-0469(1998)055<2284:AFMFTN>2.0.CO;2
Clark, T. L. (1976). Use of log-normal distributions for numerical calculations of condensation and collection. Journal of the Atmospheric Sciences, 33(5), 810–821. https://doi.org/10.1175/1520-0469(1976)033<0810:UOLNDF>2.0.CO;2
Cohard, J.-M., & Pinty, J.-P. (2000). A comprehensive two-moment warm microphysical bulk scheme. I: Description and tests. Quarterly Journal of the Royal Meteorological Society, 126(566), 1815–1842. https://doi.org/10.1256/smsqj.56613
Seifert, A., & Beheng, K. D. (2001). A double-moment parameterization for simulating autoconversion, accretion and selfcollection. Atmospheric Research, 59–60, 265–281. https://doi.org/10.1016/S0169-8095(01)00126-0
Swift, D. L., & Friedlander, S. . (1964). The coagulation of hydrosols by brownian motion and laminar shear flow. Journal of Colloid Science, 19(7), 621–647. https://doi.org/10.1016/0095-8522(64)90085-6