Review:
Parameterization of the collision-coalescence process using series of basis functions: COLNETv1.0.0 model development using a machine learning approach

Camilo Fernando Rodríguez-Genó, Léster Alfonso

This study introduces a parameterization of the collision-coalescence process of cloud droplets, which models the cloud droplet size distribution as a sum of two lognormal distributions. Within this parameterization, the time rates of change (tendencies) of five moments of the distribution are predicted by a deep neural network – traditionally, these moments are either computed by solving an integral or by using pre-computed lookup tables.
The new parameterization is compared against a reference solution, the explicit model developed by Bott (1998). In this comparison, the distribution moments obtained from the new parameterization deviated by less than 10% from those of the reference solution.

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?)

Specific comments

- L 9: “drop spectrum”, not “drop spectra” (it’s singular)
- L 15: “stablish” should probably be “establish”
- L 23: “who used” instead of “whom employed”
- L 24: “has shown”, not “have shown”
- 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.

- 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?
- 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.
- L 36: The abbreviation DSD has already been introduced in L 21.
- 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"
- L 44: Also mention numerical diffusion as one of the major problems with bin microphysics? See e.g. [1]
- 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.**
- L 91: This introduction to machine learning seems kind of out of place, especially after the previous paragraph already talks about deep neural networks.
- 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.
- Neural network architecture: How did you come up with this specific architecture? Did you try other (e.g., simpler) architectures as well?
- 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.

- 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?
- 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.
- 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.

Figures

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

References:

[1] Khain et al., 2000 , A., M. Ovtchinnikov, M. Pinsky, A.Pokrovsky, and H. Krugliak (2000): Notes on the state-of-the-art numerical modeling of cloud microphysics. Atmos. Res., 55,159–224, https://doi.org/10.1016/S0169-8095(00)00064-8.

[2] Hastie, T., Tibshirani, R.,, Friedman, J. (2001). The Elements of Statistical Learning. New York, NY, USA: Springer New York Inc..