The authors thank the anonymous referees for their helpful comments that improved the quality of the manuscript.

Comments from Anonymous Referee # 1 and answers from the authors

General comments
The study introduces a new parameterization of the collision-coalescence process that is based on the results from machine-learning procedures, with an aim to eventually use it in weather forecasting models. The authors utilized 100,000 size distributions of drops (including both cloud droplets and raindrops) to obtain the tendencies (time derivative) of 0th-5th moments, which were used for training a machine (80%) or evaluating the machine's predictions (20%). Each droplet size distribution was assumed to be a composite of two lognormal size distributions, represented by 6 parameters. The paper compares the evolutions of drop size distributions predicted by the machine-learning based parameterization and explicitly calculated by the method in Bott et al. (1998). The authors concluded that the differences were always less than 10% and therefore it has a promising potential for the future implementation in weather forecasting models.
The overall idea of utilizing the machine-learning method is innovative and aligns with what the cloud-modeling community has started working on in recent years. The results of the study are interesting and provide promising suggestions for the future model improvements. At the same time, the paper seems to require some improvements in its structures and also in providing sufficient information. Most importantly, the conclusions would become much more solid and significant if (i) more than one test simulation is done and/or (ii) if the comparison to an existing parameterization is shown. Regarding (i): although a large number of samples were used for training the machine, the overall evaluation of the new parameterization seems to rely only on one simulation (Table 4), particularly its comparison with the explicit calculation by Bott et al. (1998) under the same condition. The prediction accuracy must be somewhat dependent on each case and it is not known if this one test case falls in the "well-" or "badly-" predicted group. Regarding (ii): the prediction would always have some errors, but the magnitude of the errors is important, particularly in comparison to errors made by other existing parameterizations. Therefore, I think (i) more test simulations to compare the predictions with Bott's calculations and/or (ii) comparison with existing two-moment parameterization is necessary to draw a solid conclusion. I would highly suggest (ii). Detailed suggestions are listed below.

Answer: Regarding (i), 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 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.

Regarding (ii), an additional comparison has been included in the revised version of the manuscript, taking into account an extra parameterization, as suggested by the referee. The

popular WDM6 (WRF Double Moment 6-class) parameterization was used in the simulation, using the same initial conditions and simulation parameters. The results and discussion of the comparison have been included in the updated version of the manuscript. It was the intention of the authors to include a second extra parameterization in the paper ((Seifert & Beheng, 2001)), but because deadline issues and the extensive work needed, it was not included.

Specific comments

Lines 12-13: It seems very important to clarify what was calculated and what was predicted/estimated. Since it's supervised learning, the machine did not calculate the moments based on equations, but they must have been calculated in advance elsewhere and the results (inputs & output) were fed into the machine to train it. Afterwards, during the testing/validation phase, the total moments were predicted, not calculated by physical equations, by the trained machine. I understand the overall meaning but the readers may be misled that the machine can analytically solve the SCE and calculate the tendencies of the moments. But in reality, the machine simply gives the prediction based on what it learned before. Therefore, the word "predict/estimate" sounds more appropriate than "calculate".

Answer: The authors agree with the referee, and the wording of the abstract have been changed to reflect the fact that the Machine Learning model only predict the tendencies of the total moments, and does not solve the SCE itself.

Line 27: Adding a short explanation on a self-preserving form would be helpful (e.g., what it is, why it gets formed, etc.), especially if this is relevant to collision-coalescence.

Answer: The self-preserving form size distributions are analyzed in detail on (Swift & Friedlander, 1964), and is related to the preservation of the type of distribution function with time. Self-preserving distributions are relevant to collision-coalescence mainly because the evolution of the distribution functions due to this process can be expressed in this mathematical form.

Section 2: The structure of this section would become better if it's modified, so that there are 2.1 and 2.2, instead of only 2.1. In my observation, the first section in 2 (that I suggest to convert to 2.1) is dedicated to the time derivative of moments, regardless of collision-coalescence. Subsection 2.1 (that I suggest to change to 2.2) is providing the SCE. Mathematically speaking, I had hard time connecting the two, Eqs. 6 and 13, as Eq. 6 is not mentioned later in the paper, although I understood/knew them individually. Therefore, I suggest that the authors add a few sentences at the end of Section 2 to summarize the entire section.

Answer: The structure of the section have been modified to better organize the contents, rearranging the subsections as 2.1 and 2.2. The system of equations expressed in Equation 6 is transformed to its matrix form in Eq. 7. Equation 13 represents the way on which the total moment tendencies are calculated in the original parameterization (Clark, 1976), and is the definition of the components of vector **F** (right-hand side of the system of equations).

Lines 211-212: Although mentioned later, it would be better to mention here why the third moment tendency is not calculated.

Answer: An explanation is made about why the third moment order is not included, as suggested by the referee.

Figure 4: The figure would be more helpful if the authors instead provide a distribution (line or bar plots) of all the data rather than a scatter plot of every 100 data. Moreover, if the information (e.g., minimum, maximum, mean, median, etc.) can be provided separately for two lognormal distributions on Table 1, this figure can be omitted, as the information overlaps.

Answer: The authors agree with the redundancy of information between Figure 4 and Table 1. Thus, Figure 4 has been deleted from the article, and the rest of the figures have been renumbered.

Table 3: If the authors can add a column for a prediction score, that would be helpful too, if Matlab has a function to calculate prediction scores. The actual values of MSE may be difficult for the readers to assess the accuracy of the prediction. For example, in the text, MSEs on the other of 10-4 are considered to be a good performance, but could you explain this assessment in more detail? For example, above what number is considered a poor performance, and why, etc.

Answer: Since the values of the total moment tendencies are normalized (scale of 10^0), MSE values of 10^{-4} are considered a good performance. This explanation has been included in the manuscript, for more clarity in the text and interpretation of results. A column has also been included in Table 3, detailing the Correlation Indexes calculated between the output of the trained neural networks and the solution of the KCE.

Section 4: I think this section can be included as a subsection of 5.1 in the following Section 5, or even as 2.3 in Section 2.

Answer: The authors agree with the suggestion of the referee, and Section 4 has been relocated as subsection 2.3. All subsequent equations and sections have been renumbered accordingly.

Table 4: I understand that these conditions were chosen based on Clark (1976), but I think it would strengthen the argument that this case (or f1) is a good representation of the training data on which the machine was trained, if the authors mention the mean values in Table 1.

Answer: An explanation was included in the manuscript to reflect the fact that the initial conditions from Table 4 are in fact a good representation of the data used to train the neural networks.

Lines 357-358 and Figures 9 and 10: It is difficult to conclude whether the differences between what's predicted by the new parameterization and what's calculated by Bott's code are small enough or not, only from the figures. However, if you can add predicted values from other existing two-moment parameterizations (one frequently used in weather forecasting models), that would give the readers some insight; in Figure 10, for example, if another parameterization predicts 100 cm-3 at t=900s, then the new machine learning-based parameterization would be a better predicter. Furthermore, if such a comparison can be done for more than one case, the results would become much more solid and substantial.

Answer: In order to better demonstrate the accuracy of the developed parameterization, a comparison with the results from the collision-coalescence section of the WRF Double Moment 6-class parameterization (WDM6) have been established (Cohard & Pinty, 2000). However, a comparison methodology had to be developed, since both parameterizations are of different kinds, and their formulations are focused on different modelling philosophies. Despite that, the comparison showed promising results for the Machine Learning parameterization, particularly in the calculation of the individual moments of the drop spectrum. The proper figures and comments have been added to the manuscript, to incorporate those new findings from the comparison. It was the intention of the authors to compare the results with at least another parameterization (the one from (Seifert & Beheng, 2001)), but the amount of work needed to establish that comparison exceeded the available time offered by GMD, due to the extensive differences between the formulations of the parameterizations. Such work will be done in future research regarding the parameterization philosophy of series of basis functions here presented.

Table 5 and Figure 12: While the authors clearly state the percentage differences between the predictions and the explicit calculations, its physical meaning also needs a clarification. For example, what does the -8% error of M2 tendency prediction physically mean, and why could it be underestimated by the machine? Even more, for instance, how does this magnitude of errors compare to the errors made by other existing parameterizations?

Answer: The calculation of the percent errors are done taking the bin model results as reference. For example, a -8 % error of M2 tendency means that the predicted value of that specific moment is 8 % lower than the reference solution, regarding the reference solution itself. The causes of those differences are still subject of investigation. However, the comparison with one commonly used parameterization (explained in the previous answer) shows a better skill at predicting the statistical moments of the drop spectra than the added parameterization (WDM6). To reflect this, Table 5 has been modified to include the results of the extra parameterization considered.

Section 7: The authors conclude that the overall prediction accuracy was high, but additional analyses and/or a comparison with existing parameterizations seems to be necessary to draw the conclusion. Although the errors in Figure 12 remained less than 10%, how about other existing parameterizations? Would they be within 5%, or more than 50%? I think such a comparison would provide the readers more in-depth understanding and better assessments of the presented ML-based parameterization.

Answer: Same as the two previous comments. The authors understood that comparison with at least one extra parameterization was needed in order to provide a better assessment on the accuracy of the Machine Learning model.

Technical corrections

The authors thank the referee for the detailed revision of the technical details of the manuscript. All recommendations have been addressed, and we will only answer the ones that required specific comments.

Lines 70-72: As it approximates the droplet size distributions by two lognormal distributions, rather than using bins, I am not sure if "This approach simulates the explicit approach" is the accurate description. The strength of the authors' approach seems to be the time-varying parameters for the two lognormal distributions, in contrast to the conventional bulk schemes, which can be emphasized here.

Answer: As noted by the referee, the strength of the presented parameterization resides in the time-varying parameters for the distributions. However, is the authors' opinion that this approach could be considered a middle point between bin and bulk models, as it covers the entire size spectrum with continuous, non-truncated, distribution functions. However, we have followed the recommendations of the referee of emphasizing the main characteristic of the parameterization.

Figure 7: Since the values from the explicit calculations are the "goal/right" values, I think they should be plotted on the y axis rather than on x (i.e., suggest swapping x and y axes).

Also, the plots would look better if the x- and y ranges are identical within each plot (e.g., the plots for M1 and M4 seem to have different ranges for x and y axes).

Answer: The values from the Neural Network model are plotted in the y axis to achieve consistency across all figures in the manuscript. Since all results from the parameterization are plotted in the y axis, the authors consider that Figure 7 (renumbered Figure 6 in the revised manuscript) should not be the exception.

Regarding the ranges of the axles, while it is true that the plots would look better if the axles were identical, it is necessary to reflect that each moment has different ranges according to their characteristics. Since the values of the moments' rates are not normalized, the axles cannot be in the identical for all plots in Figure 7.

Figure 11: Though this is a small point, it would be better for the two panel plots to be placed top-and-bottom instead of left-and-right, as they share the x axis.
Answer: Following the same logic of the referee, it was the first intention of the authors to place the figure in the indicated way, prior to submission to the journal. However, after reviewing the manuscript, we noted that that configuration caused the plots to be deformed and the results could not be easily interpreted, so we opted for a left-and-right configuration of the panels.

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
1/5
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"

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. The entire paragraph has been restructured and now makes more sense for the reader.

• 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: A brief explanation of the Knudsen number and its implications on the problem at hand has been added to the introduction.

• 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: The second sentence was confusing. As both sentences shared the same references, the one referring to the geometric standard deviation has been deleted, and the remaining sentence has been properly referenced.

• 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 not to 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.

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, and remaining figures have been renumbered accordingly.
- An extra parameterization has been added (WDM6), in order to compare its results with those of the Machine Learning parameterization and the explicit model. The code for this new model has been properly referenced and included in the Code Availability section of the manuscript.
- 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.

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