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

Overall, the paper has been significantly improved in terms of both science and writing. It is clear that the authors have taken time to take the suggestions from the reviewers into account. The additional comparison with the commonly-used parameterization P-CP2000 enables readers to objectively evaluate the predictions by the new parameterization P-DNN. The authors have clarified in their response and in the text that the case in Table 4 represents the training data well. They also emphasized that the objective of this paper is an introduction of the new parameterization and its development through machine-learning. In terms of writing, however, there are still more improvements necessary before publication so that the scientific contents can be more easily conveyed to readers. Therefore, I suggest minor revisions of the paper, mostly due to the technical corrections on writing.

Specific comments

Naming:
From Chapter 4 and onwards, the authors often use "parameterized/parameterization model" and "reference/explicit model/solution" to refer to P-DNN and KCE, respectively. I suggest the authors to use consistent names (e.g., P-DNN and KCE) throughout the paper so that the readers do not get confused. The authors should first introduce the naming at the beginning of Chapter 4 and then solely use it for the rest of the paper.

Answer: The authors share the referee's concerns about naming. The manuscript has been modified to reflect this, and to be more concise and clear when referring to the different models mentioned in the paper.

Objective of Chapter 4:
It is important to clarify the meaning and objectives of the experiments done in Chapters 4 and 5. Up to Chapter 3, the detailed introductions to the equations and methodology are given. In the current form of presentation, the readers may take the comparisons done in Chapters 4 and 5 as an overall evaluation of the new parameterization P-DNN, which is solely based on one representative case. However, in reality, the comparison simply serves as an experiment on how P-DNN predicts the drop size distributions on an example case – the overall evaluation of PDNN is already done on Table 3 in Chapter 3 rather than in Chapters 4 and 5. While I understand it now, I thought the comparisons in Chapters 4 and 5 served to evaluate P-DNN, and this is why I questioned why there was only one case study in the previous round of review. Therefore, I suggest that the authors add a few sentences or a paragraph at the beginning of Chapter 4 on why this experiment/comparison is done. This will allow the readers to interpret the results just as an example, rather than misunderstand that the results (e.g., within 10 % difference from KCE) apply to all the simulations by P-DNN with any initial conditions.

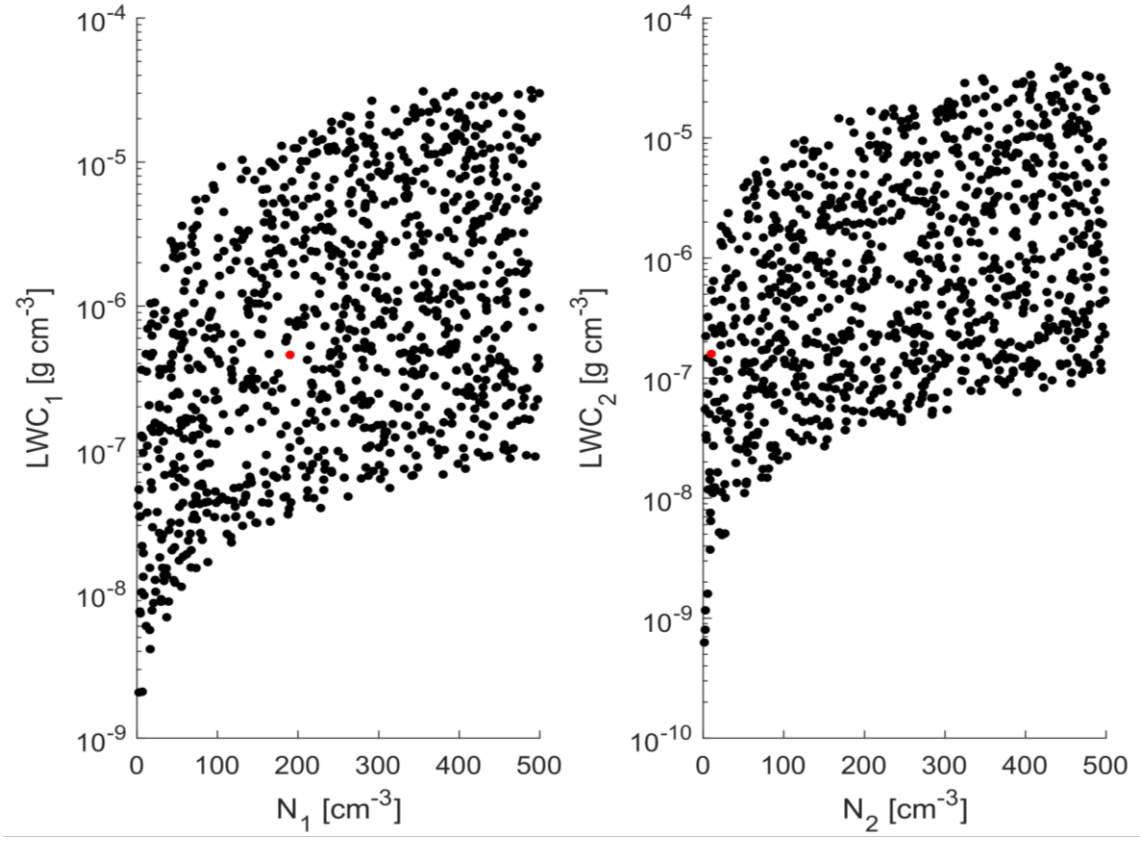
Answer: The authors acknowledge that this comment is very important to the clarity of the manuscript. Thus, an introduction has been added to Chapter 4, clarifying the interpretation and objective of the experiments, to reflect that under no circumstance the experiment should be interpreted as an overall evaluation of the developed parameterization, as this has been already done in Chapter 3, but merely an illustrating example of a simulation case.

Structure of Chapter 4:
If there is only one subchapter 4.1, it can be written as a paragraph in 4.

Answer: Following the addition of an explanatory introduction as result of the previous comment, the structure of Chapter 4 changed to reflect this. It now have sections 4.1 (experiment design and initial conditions) and 4.2 (WDM6 parameterization).

Figure 4:
If the authors can add one colored dot in each of these panel plots to show where the experiment case in Table 4 stands in this figure, that would be helpful for the readers to recognize how representative that case is.

Answer: Good point. Figure 4 has been modified to add these dots. It now looks like this:



Also, the caption also reflects what those red dots are.

Figure 7:
The simulation result with P-CP2000 is clearly different from the others, but why is it showing the existence of droplets that are smaller than the initial distribution (i.e., r < 6 um)? Is it because the size distribution was not physically calculated but was simply diagnosed by the prognosed distribution parameters? Please add some explanation.

Answer: This is caused by the fixed distribution parameters employed in its formulation. The slope parameter of the gamma distribution is determined by an analytical expression and evolves with time within certain limits, but the parameters related to the spectral breadth (width) are held fixed, thus the weird width of P-CP2000's initial spectrum. I am happy to answer any question related to P-CP2000, and also additional and more detailed information about the model can be found in Cohard and Pinty (2000).

Technical corrections

**Make sure that the references are out of the brackets when necessary (e.g., line 60).*
**Future and present tenses are mixed, please check the consistency.*
**In addition to the suggestions below, I highly suggest the thorough check on the overall writing*
by the authors.

Answer: The authors did a complete review of the manuscript, and fixed all issues related to references in brackets (and not in brackets), and the tense used throughout the paper. Also, all specific technical suggestions have been fixed, along with some others, generally improving the clarity of the manuscript.

Comments from Anonymous Referee # 2 and answers from the authors

While I do see and appreciate the effort that the authors put into revising the manuscript (namely, into the addition of the WDM6 parameterization for comparison, the restructuring of theintroduction, and the addressing of many minor reviewer comments), the major concerns that constitute my reasons for rejection are still:

1) Advantage of using a machine learning approach The advantage / added value of using a machine learning model for the task of predicting the moment tendencies (compared to e.g. a numerical quadrature) is not demonstrated. I am not sayingthat this value doesn't exist, but that it isn't established in the paper, and I think that it would have been important to do so. In their response, the authors write that the utility of the machine learningparameterization lies in a "more straightforward way of computing the moment tendencies", and that "the numerical solution of eq. (13) is a complex task, particularly the selection andimplementation of an efficient numerical methodor quadrature for the solution of double integrals". However, it seems that this complex task has been solved successfully by the authors, as they use a quadrature method to generate the output(target values) used to train the model. Given the demonstrated feasibility of numerically integrating the double integrals, I think that

simply referring to this as "a complex task" and advertising themachine learning approach as "a more straightforward way" of computing (or more precisely, predicting) the moment tendencies is not sufficient. It is probably the case that the machine learningapproach is faster than the quadrature, and that would be a good argument for why its use is advantageous, but this argument would have to be established quantitatively. I did not mean tosuggest that "an exhaustive computational or hardware-focused analysis of the problem" should be done, but if the neural network lowers the computational cost in a significant way, it should bepossible to illustrate this somehow (even if of course the exact numbers and details will depend on computational architecture), e.g. with runtime measurements for an example simulation and/orsome back-of-the-envelope calculations (where computational gains from parallelizing the predictions of the neural network could be factored in linearly). Whatever the main advantage of the neural network presented here is compared to the "competitor approaches" (it doesn't have to be computational efficiency - maybe the neural network is moreeasily adaptable to different kernels than a quadrature method?), it should be motivated / explained / demonstrated in order to make it clear that it is not a case of "machine learning for the sake of machine learning".

Answer:

Neural networks give us a better way to estimate the values of the integral (13) in the manuscript. If the parameterization was implemented from real time calculations of the integral (13) by the trapezoidal rule every time it was necessary, it would be extremely slow. The neural networks of course do not replace the computation of integrals, but since they have the ability to learn and model complex non-linear functions, they allow us (once trained) to estimate them efficiently for values of the parameters (N_1, μ_1 , σ_1, N_2, μ_2 and σ_2), for which it has not been previously calculated.

Before the widespread adoption of machine learning, the alternative previously used by other authors (Clark, 1976; Clark and Hall, 1983; Feingold et al., 1998) were the lookup tables, that are tables that stores a list of predefined values (the moment tendencies in this case). Then, in the context of our work, the lookup table is a mapping function that relates the parameters of the basis functions (N_1, μ_1 , σ_1, N_2, μ_2 and σ_2), with the total moment tendencies $\left(\frac{d\mathbf{N}\overline{\mathbf{R^p}}}{\mathbf{dt}}\right)$.

However, usually, functions computed from lookup tables have a limited domain. For larger problems, the memory and the time required to access the data increase substantially. Furthermore, preferably, we need functions whose domain is a set with contiguous values. Additionally, every time we need to calculate the integral (13), a search algorithm must be executed in order to retrieve the moment tendency for a given set of parameters, and some kind of interpolation will be needed to compute moment tendencies for values of the parameters for which it has not been calculated.

The advantage of the neural networks is that all the computational effort is dedicated to the training phase. Once we trained the networks, they replace the lookup tables and are able to map efficiently the parameters of the basis functions with total moment tendencies. A significant speed up is expected since we just need to evaluate the input parameters, and there is no need to execute a searching algorithm in order to retrieve the desired information.

This explanation has been added at the end of Chapter 3, with the objective of providing more clarity about this topic.

2) Clarity of Language / Style / Grammar: While the linguistic clarity and correctness of the revised manuscript are clearly improved compared to the previous version, there are still numerous errors (grammar and word choice) and sentences that lack in clarity.

Answer: The authors agree with the referee that a review of the entire manuscript was in order. This has been done, with the comments of both referees as guide, and the clarity of the paper has improved. Next you will find a couple of comments that we believed worth it:

- *L69 "simulates the explicit approach" –> unclear what this means*

  Answer: What we tried to say with this is that the parameterization developed by Clark (1976) and Clark and Hall (1983) differs from the traditional bulk methods in the variables it calculates. The bulk methods usually follow the evolution of one, two or more recently three selected moments of a distribution function (usually gamma), while Clark's approach follows the evolution of the parameters of those distributions (lognormals), and not the actual values of the moments per se. Thus, it "simulates" the bin approach in the way that the DSD can be easily reproduced from those parameters. Therefore, Clark follows the DSD through the explicit calculation of the distributions' parameters, and not the moments.

  Of course, the formulation of the rates of the moments is included in the system of equations, but this is incidental, and needed to close the system, not because the calculation of the moments is the objective of the parameterization. That is why we called the Clark's parameterization a "hybrid approach" to modelling cloud microphysics.

- *L 94 The abbreviation "ML" has not been defined*

  Answer: True. The ML abbreviation has been defined in the first mention of Machine Learning in the introduction.

3) Evaluation of the new parameterization for several different experiments. This was also suggested by the other reviewer but considered outside of the objective / scope of the study by the authors. That is fair, but I still think it would have made the study more solid.

Answer:

We believe that clarity is responsibility of the authors, since the following was not clearly stated in the manuscript or in the previous answers. It is important to clarify the meaning and objectives of the experiments done in Chapter 3 and 4. The overall evaluation of the novel components (ML approach) is done in Chapter 3, through the information in Table 3 and in Figures 5 and 6, and their related explanations and analyses in the text. The comparison introduced in Chapter 4, and results showed in Chapter 5, simply serve as an illustrating experiment on how the developed model predicts the DSD and bulk variables, on an example basis. Under no circumstance the experiments from Chapters 4 and 5 should be interpreted as an overall evaluation of P-DNN. An introduction has been added to Chapter 4 clearly explaining all this. That is why the authors considered the realization of more experiments outside of the scope of the manuscript.

References

Clark, T. L.: Use of Log-Normal Distributions for Numerical Calculations of Condensation and Collection, J. Atmos. Sci., 33(5), 810–821, doi:10.1175/1520-0469(1976)033<0810:UOLNDF>2.0.CO;2, 1976.

Clark, T. L. and Hall, W. D.: A Cloud Physical Parameterization Method Using Movable Basis Functions: Stochastic Coalescence Parcel Calculations, J. Atmos. Sci., 40(7), 1709–1728, doi:10.1175/1520-0469(1983)040<1709:ACPPMU>2.0.CO;2, 1983.

Cohard, J.-M. and Pinty, J.-P.: A comprehensive two-moment warm microphysical bulk scheme. I: Description and tests, Q. J. R. Meteorol. Soc., 126(566), 1815–1842, doi:10.1256/smsqj.56613, 2000.

Feingold, G., Walko, R. L., Stevens, B. and Cotton, W. R.: Simulations of marine stratocumulus using a new microphysical parameterization scheme, Atmos. Res., 47–48, 505–528, doi:10.1016/S0169-8095(98)00058-1, 1998.