Response to the reviewers: Analysis of the cloud fraction adjustment to aerosols and its dependence on meteorological controls using explainable machine learning # EGUSPHERE-2023-1667

Yichen Jia [1, 2], Hendrik Andersen [1, 2], and Jan Cermak [1, 2]

[1]Karlsruhe Institute of Technology (KIT), Institute of Meteorology and Climate Research, Karlsruhe, Germany
[2]Karlsruhe Institute of Technology (KIT), Institute of Photogrammetry and Remote Sensing, Karlsruhe, Germany

Correspondence: Yichen Jia (yichen.jia@kit.edu)

We would like to thank the two anonymous referees for their reviews of the revised manuscript and their insightful feedback. Below, the reviewers' comments and suggestions are incorporated in italics and addressed hereafter, and the authors' responses are coloured in blue. Unless otherwise stated, line numbers in this document refer to the manuscript after the first-round review (before the updates following in this response letter).

Referee 1

General comments

I appreciate the effort and work the authors have done to address my questions and issues raised in the last round. The manuscript has significantly improved and almost is ready for publication,

Thank you for the kind words and the positive evaluation of the manuscript.

Technical issues

1. line 326, two "?" I do not know what they mean.

2. line 339, same issue. Please double-check all references and format, etc.

Thank you for catching that. This issue due to a wrong citation key has been corrected throughout the manuscript.

Referee 2

General comments

On the point of justifying SHAP values of a complicated tree model, the authors make the point that basically everything statistical modeling technique is doing similar things. While that to some extent is true, it is besides the point. The point is that using extremely complicated models, that is beyond say individual scientists' grasp, to 'understand' behavior of low clouds is questionable. The understanding in this context comes from calculating SHAP values of a highly nonlinear model. While SHAP values are popular, they have their own limitations: the calculation itself depends on the software package, it assumes feature *independence, and most importantly these calculations are not meant for causal inference. That is precisely I made the point of checking the 'physical sense' of model behavior reported here. I used figure 1a as an example. As N_d increases, the SHAP values continue to scale with N_d (or lnNd). In other words, lnNd contributes to increase low cloud fraction linearly. This is not physical and not what we observe in nature. If anything, multiple lines of evidence suggest that N_d tend to saturate after a rather low threshold (Rosenfeld, et al., 2012; 2019; Yuan et al., 2023). This behavior is determined by low cloud precipitation formation and the threshold is quantified to be around 60-80 per CC. Similarly, I mentioned systematic biases associated with N_d calculations and low cloud fraction. This is also physical and built in with the data the authors used. In other words, there is a strong component of covariance between these two that does not have physical implications but artificial correlation. These are first order considerations when attempting to 'explain' low cloud fractions. The standardization procedure is also puzzling since it means that all the quantities calculated here are based on local data distribution and thus not really comparable between different 5x5 grids. It is however fine for the authors to try to analyze the data and report their statistical findings. It may be a useful endeavor for readers of this paper. However, it is clear to me that the authors do not seem to have a solid background in the physical understanding of low cloud formation and aerosol effects on them while attempting to extract physical explanations from statistical models. To this point, there is a saying by Einstein that says 'Everything should be made as simple as possible, but not simpler'. The authors complied a long list of variables and used them to best fit the data. I suggest the authors to write the paper based on actual data and do not make too much into causal inference and discuss physical implications because the tool used here does not allow that. Particularly, the results do not always make physical sense anyway*

We thank the reviewer for her/his input, and see three main points that the reviewer is concerned about: 1) model complexity, SHAP limitations, biases in the data, 2) potentially unphysical relationships, and 3) standardization procedure. In the following, we address these three concerns.

1. While in the original version of the manuscript, we already discussed the potential issues related to retrieval biases leading to positive N_d–CLF correlations in the data section, we take the concerns of the reviewer seriously, and have added a new section (Sect. 2.3.3, titled "Limitations of observation-based machine-learning of aerosol-cloud processes") dedicated to discussing the limitations of this study regarding the data sets and methods used. Within this section, we discuss the general limitation of the SHAP approach, the built-in limitation of the data sets, the somewhat limited interpretability of the method with regard to physical processes, and the appropriate interpretation of sensitivities in this study. Many of these limitations are not limited to this study, though, but rather general limitations for observation-based machine-learning of aerosol-cloud processes (or more specifically the cloud fraction adjustment). In this section, we now explicitly state that all these limitations should be considered when interpreting the results. To be more careful with the interpretation of the results, we have also removed "in a physically meaningful way" from the last sentence in the fourth point of the conclusion section. Another outlook point has been included at the end of the manuscript: "In addition, ... processes."

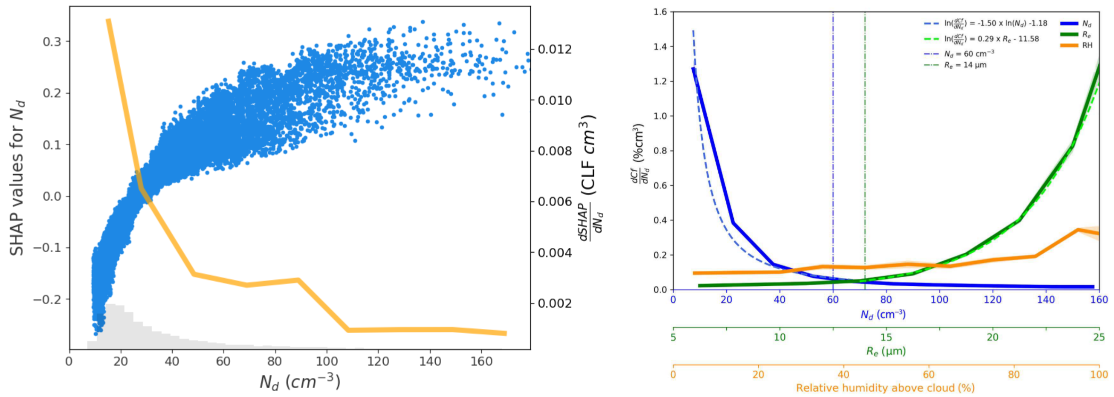

Figure 1. The left panel is the same as Figure 1 (a) in the manuscript. The orange line represents the rate of change in SHAP values for N_d due to the N_d change (dSHAP/dN_d) across binned N_d values. The right panel is taken from Yuan et al. (2023). The solid blue line shows the rate of CLF change with respect to N_d change (dCF/dN_d) as a function of binned N_d. Here we highlight the similarity between the orange line in the left panel and the blue line in the right panel.

2. Unphysical relationships: The reviewer states that "As N_d increases, the SHAP values continue to scale with N_d (or lnN_d)", and later mentions that the effect of N_d on CLF saturates at low N_d values ($\sim$ 60–80 per CC), as e.g. found in Yuan et al. (2023). We agree to the latter statement. However, it is important to point to the difference between the CLF–N_d and CLF–lnN_d relationships. In Yuan et al. (2023), the N_d–CLF relationship (dCF/dN_d) is shown to be sensitive at very low N_d values and then approaches 0 in the value range mentioned by the reviewer. We see the same behaviour in our data sets (see Fig. 1 in this response), which is why we use the natural logarithm of N_d to create a linear relationship. We have added the left panel of Fig. 1 of this response as an additional new panel to the first figure of the manuscript to show this difference between the SHAP dependence plots for N_d and lnN_d. As shown in the previous version of the manuscript, lnN_d SHAP values continue to scale linearly with lnN_d, while the increase in N_d SHAP values due to increasing N_d saturates at higher N_d values (for this specific geographical window, the sensitivity approaches 0 at about 100 per CC). This behaviour holds for not only this specific example geographical window, but also for others in general. Overall, the development of the CLF sensitivity to N_d as a function of N_d behaves remarkably similar when compared to Yuan et al. (2023) (see Fig. 1 in this response for a direct comparison). This implies that the use of SHAP values in explaining XGB captures the physical relationships similar to the study mentioned by the reviewer. We thus reject the notion of the reviewer that the central physical relationships discussed in our manuscript are unplausible, but rather that the comparison to the study mentioned by the reviewer supports our findings. We do see though that the communication of these results was not ideal but believe that the new panel in Fig. 1 in the manuscript helps the reader better understand this important aspect of our study. Section 2.3.2 has also been revised accordingly.

3. The z-score transformation of data sets is a typical procedure for studies also aiming to compare cloud responses to multiple different meteorological variables (e.g. Scott et al., 2020; Andersen et al., 2023). The advantage of this prepro-

cessing step is that it places all meteorological variables on the same scale allowing a comparison to a typical variation in each meteorological field. We acknowledge that the downside of this approach, i.e. that standard deviations change regionally, making the interpretation of the spatial patterns more difficult, and that readers may be interested in the sensitivities from the non-standardized data. To address this, we have added a supplementary material showing the results from non-standardized data. We can clearly see that, in the supplement, the example SHAP dependence plots and the geographical patterns of sensitivities and IAIs are nearly identical to the corresponding ones in the manuscript, only absolute magnitudes are notably different as expected. The last part of Sect. 2.1 has been rewritten and expanded. Relevant discussions have also been added in Sect. 2.3.2.

Minor modifications independent of the reviewer comments

"the" is added before "cloud fraction" in the title.

Space between "ln" and "N_d" is removed throughout the text of the manuscript (ln N_d to lnN_d).

P. 5 L. 130: "(see Sect. 2.3.2 in detail)" has been added after "logarithm of N_d is taken".

Figure 5 caption: "are" to "is".

Figure 6 caption: "colourbar" to "colourbars".

P. 17 L. 386: "Including many confounding and influencing factors as a whole," has been added before "The explainable machine learning technique ...".

P. 17 L. 389: we have added a sentence "The statistical sensitivities ... physical understanding of the system." before "The main findings of ... as follows:".

References

Andersen, H., Cermak, J., Douglas, A., Myers, T. A., Nowack, P., Stier, P., Wall, C. J., and Wilson Kemsley, S.: Sensitivities of cloud radiative effects to large-scale meteorology and aerosols from global observations, Atmospheric Chemistry and Physics, 23, 10 775–10 794, https://doi.org/10.5194/acp-23-10775-2023, 2023.

Scott, R. C., Myers, T. A., Norris, J. R., Zelinka, M. D., Klein, S. A., Sun, M., and Doelling, D. R.: Observed Sensitivity of Low-Cloud Radiative Effects to Meteorological Perturbations over the Global Oceans, Journal of Climate, 33, 7717–7734, https://doi.org/https://doi.org/10.1175/JCLI-D-19-1028.1, 2020.

Yuan, T., Song, H., Wood, R., Oreopoulos, L., Platnick, S., Wang, C., Yu, H., Meyer, K., and Wilcox, E.: Observational evidence of strong forcing from aerosol effect on low cloud coverage, Science Advances, 9, eadh7716, https://doi.org/10.1126/sciadv.adh7716, 2023.