Response to interactive comment of anonymous referee 1 —

Julia Fuchs[1, 2], Jan Cermak[1, 2], and Hendrik Andersen[1, 2]

[1]Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany.

[2]Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany.

contact: julia.fuchs@kit.edu

"Building a cloud in the Southeast Atlantic: Understanding low-cloud controls based on satellite observations with machine learning" by Fuchs et al. applies a machine- learning program to satellite observations and studies the factors that influence cloud properties in the southeast Atlantic. The method is novel and by itself worthy of publication. The findings on sub-regional variability in dominant factors are interesting and promote better understanding of the climate in the region. The manuscript is written well. I recommend publication. The authors may consider the following suggestions.

General Comments:

Discussion on the data size and the robustness of statistics would be helpful. The variables and their spatial and temporal ranges are given in Section 2.1 and Section 2.2. But I find it difficult to determine whether some sharp features (e.g., in Figure 3d around 282.7K) are a result of poor counting statistics.

The GBRT models are computed based on approximately 2000 data points per parameter (now added on p.3, l.19). A robust performance of these models is shown in terms of the R2 (NRMSE), which presents a good agreement of the predicted vs. observed cloud property based on 10 model runs of an independent (unseen) dataset. The robustness of the model toward overfitting to the training dataset is ensured by the cross-validated tuning of the hyperparameter, the choice of the robust Huber loss function and the implementation of an early stopping rule. The section of the manuscript (p.4, l.9) is modified for clarity.

The sharp feature observed in Figure 3d for the NE subregion is shown in Fig. 1 together with a 2-dimensional frequency plot of the total data counts and the data mean per T700 bin. A good agreement between modeled and observed relationship is shown, and the sharp feature is associated with sufficient data. Thus, this case shows how the model is able to capture the data inherent relationships. However, the marked steps in the partial dependencies (e.g. Fig. 5) are most likely artifacts due to the decision tree based algorithm. This aspect is added on p.5, l.2: "Marked steps in the partial dependencies have to be interpreted with caution (e.g. Fig. 5), as they can be in part caused by the decision tree based algorithm, dividing the parameter space into separate regions.".

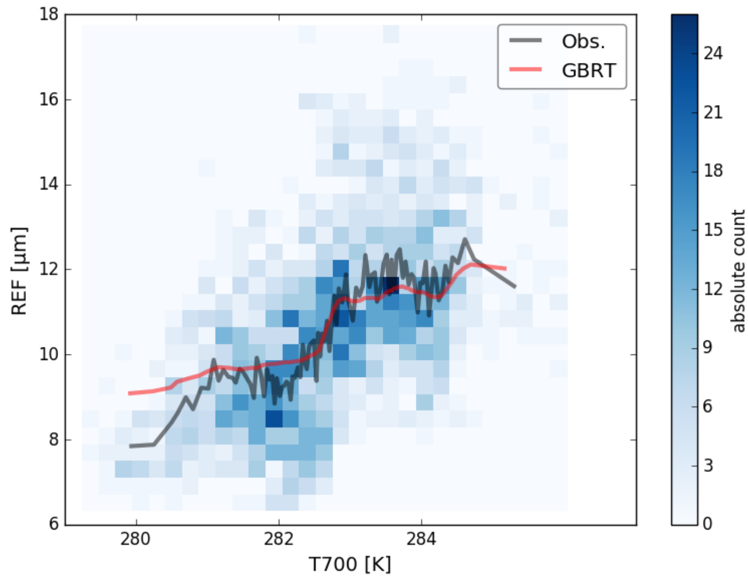

Figure 1: Predicted (GBRTs; red) versus observed (Obs.; black) mean REF binned to 98 T700 percentiles (1st - 99th) of the observation data for the NE subregion. Two-dimensional absolute frequencies of observations colored in blue.

Detailed Comments:

Page 1, line 20. Remove the first comma.

Done.

Page 3, line 33. What is meant by "generalize, its performance and computational demand"?

The ability of a GBRT model to generalize means that the model is capable to predict an output with good agreement to the observations (R2, NRMSE) based on an unseen dataset. The more the model learns (without overfitting to the dataset) the better it is able to predict (performance), however, the longer is the training and running time for the model to be computed. The sentence is rephrased as follows: "In general, a high number of boosting iterations and a low learning rate will increase the models ability to make predictions on an unseen dataset (generalize), its performance and computational demand during training."(p.4, l.1)

Page 5, line 12. Rephrase "relative humidity is essential for cloud formation processes and characteristics".

The sentence is rephrased as follows: "As free tropospheric and cloud-level humidity influence dry-air entrainment and cloud characteristics in marine low clouds (Wood, 2012; Jones et al., 2014; Bretherton et al., 2013; Andersen et al., 2017), relative humidity values at 700, 850 and 950 hPa are selected as predictors." (p.5, l.21)

Page 6, line 15. Break down the long sentence.

The sentence is broken down as follows: "The application of the GBRTs aims at finding subregional patterns of relevant low-cloud drivers, without creating a model which fully covers the interactions between clouds and their environmental conditions. The predictor set was selected in a way to reduce covariation. Thus, the choice of predictors reflects the compromise between characterizing the atmospheric state sufficiently without creating a model that lacks interpretability."(p.6, l.25)

Page 7, line 8. "LTS is most sensitive to CF". Did you mean "CF is most sensitive to LTS"?

Yes, thanks for this comment. The sentence is modified accordingly.

Page 10, line 10. "the reduction of CF by subsiding dry air". Isn't subsidence usually associated with higher stability and more clouds?

Yes, however, a study by Myers and Norris (2013) showed further that subsidence can also reduce cloudiness for the same value of LTS, which is explained by a lowering of the marine boundary layer. The reference is added to the manuscript (p.9, l.32).

Page 10, the paragraph starting in line 27, or later. Figures 5-11 are from only one model run selected at random. How representative are these snapshots of all model runs?

The two-variable partial dependencies are essentially the same as the one-variable partial dependencies only for two predictors. Thus, the same range between maximum and minimum of the one-variable partial dependence obtained from all model runs (shaded area in e.g. Fig. 3) is expected for the two-variable partial dependencies. This is now mentioned in the caption of Fig. 5: "For this illustration only one model run is selected at random as it represents all model runs with error ranges comparable to that of the one-variable partial dependencies." Figure 2 shows similar patterns obtained from three different model runs.

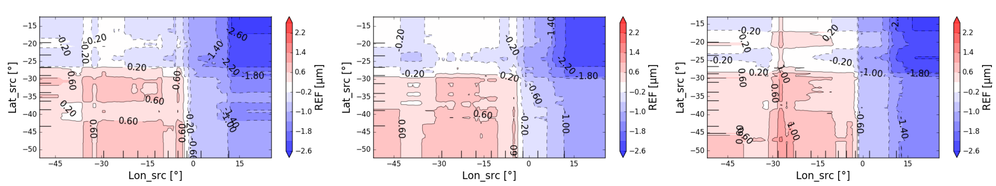

Figure 2: Two-variable partial dependence of REF on Lon_src and Lat_src in the in the SW subregion. The three panels show three SW model runs selected at random.

Page 11, line 1. Remove the first comma.

Done.

Page 11, line 3. Remove the first comma.

Done.

References

Andersen, H., Cermak, J., Fuchs, J., Knutti, R., and Lohmann, U. (2017). Understanding the drivers of marine liquid-water cloud occurrence and properties with global observations using neural networks. *Atmospheric Chemistry and Physics*, 17(15):9535–9546.

Bretherton, C. S., Blossey, P. N., and Jones, C. R. (2013). Mechanisms of marine low cloud sensitivity to idealized climate perturbations: A single-LES exploration extending the CGILS cases. *Journal of Advances in Modeling Earth Systems*, 5(2):316–337.

Jones, C. R., Bretherton, C. S., and Blossey, P. N. (2014). Fast stratocumulus time scale in mixed layer model and large eddy simulation. *Journal of Advances in Modeling Earth Systems*, (6):206–222.

Myers, T. A. and Norris, J. R. (2013). Observational evidence that enhanced subsidence reduces subtropical marine boundary layer cloudiness. *Journal of Climate*, 26(19):7507–7524.

Wood, R. (2012). Stratocumulus Clouds. *Monthly Weather Review*, 140(8):2373–2423.