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.".

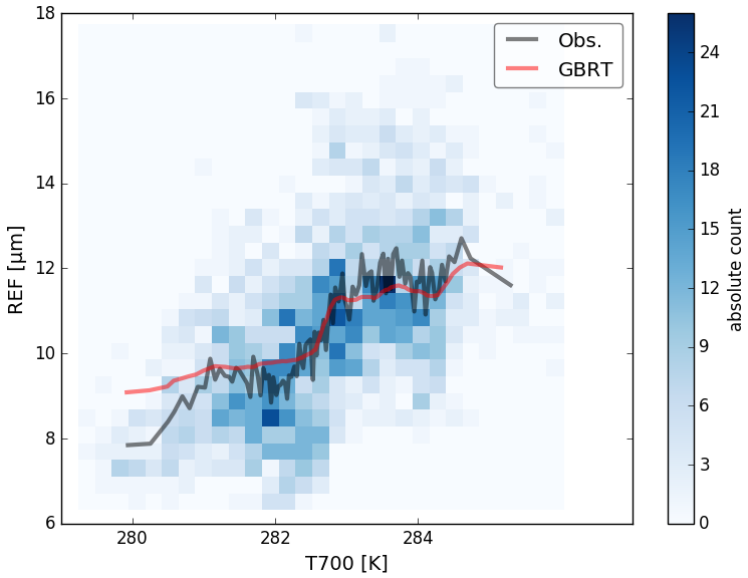

Figure 1: Predicted (GBRTs; red) versus observed (Obs.; black) mean REF binned to 98 T700 percentiles (1st - 99th) of the observation data for the NE subregion. Two-dimensional absolute frequencies of observations colored in blue.

Detailed Comments:

Page 1, line 20. Remove the first comma.

Done.

Page 3, line 33. What is meant by "generalize, its performance and computa-

tional 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.

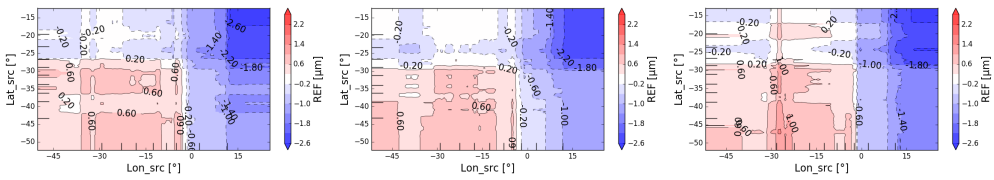

Figure 2: Two-variable partial dependence of REF on Lon_src and Lat_src in the in the SW subregion. The three panels show three SW model runs selected at random.

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.

Response to interactive comment of anonymous referee 2 —

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

This manuscript disentangles aerosol effects on the southeast Atlantic stratocumuls deck from meteorological effects through the use of a machine learning approach labeled Gradient Boosting Regression Trees (GBRTs). It is welcome to see a recognition of both impacts, and the use of an innovate approach to discriminate them. The use of lat_src and lon_src is nice. The results are sensible. I do however feel the study suffers from over-interpretation. One concern is the focus on only the cloud fraction and the cloud effective radius (REF) as the cloud properties. While the REF is influenced by aerosol, it is also a function of the liquid water path. A more straightforward physical relationship is that between AOD (CCN) and the cloud droplet number concentration (Nd), which can be estimated as a function of REF and the cloud optical depth. Cloud deepening is likewise better interpreted through the use of LWP than of REF. Another concern is the lumping of July-August-September. It is by now well appreciated that the biomass-burning aerosol is more likely to be present within the boundary layer in July, moving up in altitude through September, when it is more likely to be above the cloudy boundary layer. Different cloud responses would be anticipated as a function of the month. A useful additional analysis is to examine the GBRT results as a function of month, and interpret them as a function of the varying cloud-aerosol vertical structure.

The study was designed to focus on cloud fraction and cloud effective radius in order to test and interpret the GBRT models on one relevant micro- and one relevant macrophysical cloud property during the biomass-burning season. Using the cloud droplet number concentration is appreciated, however, as it is derived from COT and REF, and based on assumptions on cloud vertical profile, additional uncertainties would be introduced (Grosvenor et al., 2018). We chose to avoid this, because we try to capture the cloud system as completely as possible with the statistical model. As such, we include information

on factors that also determine LWP. As variability among these LWP-predictors is simulated in the computation of the sensitivities, we thereby indirectly constrain LWP effects on REF. To account for the referee's suggestion LWP as an essential cloud property is analyzed and results support the interpretation of cloud thickening under stable conditions in all subregions (Fig. 1a), as well as during westerly disturbances, especially in the western subregions (Fig. 1b). In the manuscript the LWP-effects refer to outcomes of a comparable study by Fuchs et al. (2017) where LWP is discussed and 'self-constraining model' is now detailed more clearly.

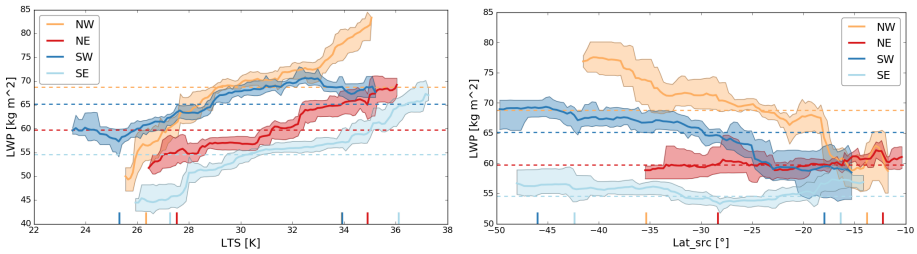

Figure 1: Mean partial dependence of LWP on LTS (left) and source latitude of air mass (right) in the four subregions (colors).

The aggregation of the months July-August-September (JAS) was conducted for better comparability to previous studies (Painemal et al., 2014; Andersen and Cermak, 2015; Adebiyi and Zuidema, 2018) investigating the same season. While we agree with referee 2, changes of the aerosol and boundary layer occur on all scales, so that the assumptions outlined by referee 2 need to be made independent of scale. The intraseasonal variability contained in the training of the GBRT model contribute to the relationships during the investigated season and must be taken into account for the interpretation of results. For this reason, we have now computed monthly GBRT models and included the results concerning the aerosol-cloud relationships in the manuscript. The following figure and text are added to the manuscript. "Figure 7 shows AOD-REF

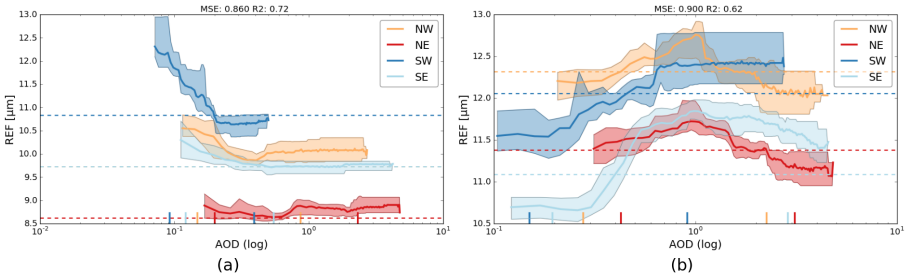

Figure 2: Mean partial dependence of REF on AOD in the four subregions (colors) in July (a) and September (b).

partial dependencies for the months of July and September separately. While during July, REF seems to decrease with increasing AOD, especially in the SW subregion, during September the opposite relationship is found. The contrasting relationships may be related to differences in the vertical distribution of aerosols and clouds in the Southeast Atlantic. During July, aerosol and cloud layers are frequently entangled, facilitating ACI, whereas in September they can be well separated (Adebiyi et al. 2018). During this time, absorbing aerosol may increase the stability and trap humidity in the boundary layer, potentially leading to the observed relationship. The JAS partial dependence between AOD and REF can thus be viewed as a summary of these patterns. However, it is not the study's focus to separate the different aerosol effects mentioned earlier, but to analyze the overall influence of aerosols on clouds during the biomass-burning season."(p.14, l.10)

Other comments follow:

1. I am not completely comfortable with the use of the 8-day MODIS L3 product used as opposed to shorter time scale, as the 8-day time scale will average over the synoptic time scale and is far longer than the cloud adjustment

time scale of 1-2 days. The authors mention that an 8-day time scale "allows for the large-scale and thermodynamic forcings of cloud properties to be combined", but I remain unclear what this means exactly. In several places in the manuscript the authors refer to processes that occur at much smaller time scales, such as the cloud microphysical response to aerosol. Instead it seems to me the 8-day time scale is primarily capturing a portion of the monthly evolution in the aerosol-cloud vertical structure and seasonal meteorological cycle. Also, the 8-day time scale should be explicitly mentioned in the abstract. The study focuses on processes on aggregated time scales (8-day), assuming that cloud adjustments due to aerosols, though acting on smaller time scales, are detectable in the aggregated data set at the same time as changes of thermodynamic and dynamic conditions. While daily or hourly data might underestimate e.g. the effect of LTS on the cloud cover, the influence of aerosols on the cloud cover might be underestimated by the 8-day aggregation. In particular, since aerosol and cloud properties are not retrieved at the same time in a given location. Eight-day averages are taken to represent the mean states of both at that time scale. These aspects are important and now more explicitly addressed in the manuscript. "The temporal resolution of 8 days allows to combine large-scale, thermodynamic and aerosol forcings of cloud properties simultaneously on a synoptical scale. However, it must be taken into account that clouds adjust on different time scales (hours to several days) to their environment (Klein, 1997; McCoy et al., 2017; Adebiyi and Zuidema, 2018) and thus processes relevant on shorter time scales might be underrepresented in the data set."(p.3, l.8) The 8-day time scale is now introduced in the abstract (p.1, l.5).

2. an issue with using the relative humidity at 950 hPa is that changes in RH are more likely to reflect co-variations with other factors such as the

cold-temperature advection (I suspect this explains the stronger relationship between RH_950hpa and REF in the SE sub-region) and cloud-top inversion strength. Have the authors examined the cross-correlations between their predictors?

Thank you for pointing out this issue. Correlations between the predictors were examined in advance and influenced the choice of predictors to reduce the covariation. Cold-temperature advection and cloud-top inversion strength are not explicitly chosen as predictors, but are assumed to be represented in the data set by other parameters such as wind speed, sea surface temperature and LTS. The manuscript is modified on p.6, l.26: "The predictor set was selected in a way to reduce covariation."

3. how is it that the machine learning approach is able to grasp non-linear relationships? The description of the technique presented on p. 4 still seems to present it as a basically linear technique.

The GBRT algorithm is based on decision trees which are capable of representing non-linear dependencies between predictor and predictand. The parameter space is iteratively split with the goal to minimize a loss function. The sum of the linear decisions of each tree in the ensemble can then represent non-linear relationships. The manuscript is modified on p.3, l.24.

4. It is worth mentioning that the larger region encompassing the 4 subregions has been previously examined in Klein and Hartmann 1993.

This reference is added on p. 3, l.13: "In this study CF and REF are simulated based on a selected predictor set (AOD and meteorological parameters) in the SEA (10°–20° S, 0°–10° E, as analyzed in Klein and Hartmann (1993)) using Gradient Boosting Regression Trees (GBRTs)."

References

Adebiyi, A. A. and Zuidema, P. (2018). Low Cloud Cover Sensitivity to Biomass-Burning Aerosols and Meteorology over the Southeast Atlantic. *Journal of Climate*, 31(11):4329–4346.

Andersen, H. and Cermak, J. (2015). How thermodynamic environments control stratocumulus microphysics and interactions with aerosols. *Environmental Research Letters*, 10(2):024004.

Fuchs, J., Cermak, J., Andersen, H., Hollmann, R., and Schwarz, K. (2017). On the Influence of Air Mass Origin on Low-Cloud Properties in the Southeast Atlantic. *Journal of Geophysical Research: Atmospheres*, 122(20):11,076–11,091.

Grosvenor, D. P., Sourdeval, O., Zuidema, P., Ackerman, A., Alexandrov, M. D., Bennartz, R., Boers, R., Cairns, B., Chiu, J. C., Christensen, M., Deneke, H. M., Diamond, M. S., Feingold, G., Fridlind, A., Hünerbein, A., Knist, C. L., Kollias, P., Marshak, A., McCoy, D., Merk, D., Painemal, D., Rausch, J., Rosenfeld, D., Russchenberg, H., Seifert, P., Sinclair, K., Stier, P., van Diedenhoven, B., Wendisch, M., Werner, F., Wood, R., Zhang, Z., and Quaas, J. (2018). Remote sensing of droplet number concentration in warm clouds: A review of the current state of knowledge and perspectives. *Reviews of Geophysics*.

Klein, S. and Hartmann, D. (1993). The seasonal cycle of low stratiform clouds. *Journal of Climate*, 6:1587–1606.

Klein, S. A. (1997). Synoptic Variability of Low-Cloud Properties and Meteorological Parameters in the Subtropical Trade Wind Boundary Layer. *Journal of Climate*, 10(8):2018–2039.

McCoy, D. T., Eastman, R., Hartmann, D. L., and Wood, R. (2017). The change in low cloud cover in a warmed climate inferred from AIRS, MODIS, and ERA-interim. *Journal of Climate*, 30(10):3609–3620.

Painemal, D., Kato, S., and Minnis, P. (2014). Boundary layer regulation in the southeast Atlantic cloud microphysics during the biomass burning season as seen by the A-train satellite constellation. *Journal of Geophysical Research: Atmospheres*, 119(19):11,288–11,302.

Building a cloud in the Southeast Atlantic: Understanding low-cloud controls based on satellite observations with machine learning

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.

Correspondence: Julia Fuchs (julia.fuchs@kit.edu)

Abstract. Understanding the processes that determine low-cloud properties and aerosol-cloud interactions (ACI) is crucial for the estimation of their radiative effects. However, the covariation of meteorology and aerosols complicates the determination of cloud-relevant influences and the quantification of the aerosol-cloud relation.

This study identifies and analyzes sensitivities of cloud fraction and cloud droplet effective radius to their meteorological and aerosol environment in the atmospherically stable Southeast Atlantic during the biomass-burning season based on an 8-day-averaged dataset. The effect of geophysical parameters on clouds is investigated based on ~~a~~ machine learning technique, gradient boosting regression trees (GBRTs), using a combination of satellite and reanalysis data as well as trajectory modeling of air-mass origins. A comprehensive, multivariate analysis of important drivers of cloud occurrence and properties is performed and evaluated.

The statistical model reveals marked subregional differences of relevant drivers and processes determining low clouds in the Southeast Atlantic. Cloud fraction is sensitive to changes of lower tropospheric stability in the oceanic, southwestern subregion, while in the northeastern subregion it is governed mostly by surface winds. In the pristine, oceanic subregion large-scale dynamics and aerosols seem to be more important for changes of cloud droplet effective radius than in the polluted, near-shore subregion, where free tropospheric temperature is more relevant. This study suggests the necessity to consider distinct ACI regimes in cloud studies in the Southeast Atlantic.

1 Introduction

Low-level clouds play a major role in the climate system via their impact on the Earth's energy budget and water cycle (Boucher et al., 2013). However, the estimation of their potentially large negative radiative effect is prone to large uncertainties as processes that govern cloud micro- and macro-physical properties, i.a. aerosol-cloud interactions (ACI), and the impact of changing environmental conditions on low clouds ~~,~~are not sufficiently understood (Bony and Dufresne, 2005; Medeiros et al., 2008). Maritime stratocumulus clouds, persisting over the relatively clean southern oceans are thought to be especially sensitive to aerosols, exerting a strong cloud albedo effect of -0.2 W m^{-2} (Platnick and Twomey, 1994; Quaas et al., 2008). One of these regions, the Southeast Atlantic (SEA), has become a very popular region for studies of low-cloud processes and ACI in the last decade (e.g., Chand et al., 2009; Muhlbauer et al., 2014; Painemal et al., 2014; Andersen and Cermak, 2015;

Adebiyi et al., 2015; Fuchs et al., 2017).

The semi-permanent low-cloud cover of the SEA is driven by the cold Benguela current offshore the Namibian/Angolan coast and maintained by large-scale subsidence (Wood, 2012). During the biomass-burning season in July-August-September (JAS), carbonaceous aerosols are advected over the oceanic boundary layer and frequently build a thick layer above the clouds. Black carbon aerosol particles can act as cloud condensation nuclei as they are entrained at cloud-top (Seinfeld et al., 2016) or indirectly alter cloud cover through the strengthening of the inversion by absorption of shortwave radiation above the cloud (Wilcox, 2010; Bond et al., 2013; Li et al., 2013).

Despite advances on the basis of large eddy simulations (e.g., Yamaguchi and Randall, 2008; Jones et al., 2014), lagrangian approaches (e.g., Mauger and Norris, 2010) and observational studies (e.g., Zuidema et al., 2016), the complex mechanisms between low clouds, boundary layer processes, thermodynamics and large-scale circulation are not sufficiently understood. Untangling the drivers of cloud properties is challenging, as meteorological parameters and aerosols covary (Mauger and Norris, 2007; Fan et al., 2016), vary spatially and have different time scales (Jones et al., 2014; Eastman et al., 2016; de Szoeke et al., 2016).

In a recent study, Fuchs et al. (2017) showed that air-mass origins can explain some of the variability of cloud microphysics in the SEA, with clear spatial differences in the involved processes. Analyses of cloud sensitivities in the SEA would therefore benefit from a subregional determination of large-scale, thermodynamic and aerosol drivers of cloud property changes. Relevant mechanisms for changes of low-cloud properties are studied here focusing on two questions:

- What are the subregional differences of cloud sensitivities to various geophysical parameters?
- How do these determinants influence cloud properties and their response to atmospheric aerosol loading?

In this study a machine learning approach is used to predict cloud fraction and cloud droplet effective radius in the SEA based on satellite and reanalysis data. This study does not aim to simulate microphysical cloud processes and individual feedback mechanisms at the level of detail of a cloud-resolving model, but instead intends to represent non-linear patterns of cloud adjustments to the large-scale and thermodynamic environment in a coherent, multivariate statistical model.

2 Methods

2.1 Data

Cloud fraction (CF), cloud droplet effective radius (REF) and aerosol optical depth (AOD) are obtained from the 8-day level 3 (L3) product of the MODerate-resolution Imaging Spectroradiometer (MODIS) instrument aboard the Aqua platform (collection 6). The data cover a temporal range from 2002 to 2012 during the biomass-burning season in July-August-September. The REF product is based on single-layer liquid clouds to avoid effects of overlapping cirrus clouds (Hubanks et al., 2018).

The following thermodynamic and dynamic parameters of the ERA-Interim reanalysis dataset of the European Centre for Medium-Range Weather Forecasts (ECMWF) are used: lower tropospheric stability (LTS), relative humidity at 950/850/700

hPa (RH950, RH850, RH700), surface wind speed at 10 m (WSP10), sea surface temperature (SST) and temperature at 700 hPa (T700), zonal wind speeds at 600 hPa (U600) and mean sea level pressure (MSLP).

The ERA-Interim reanalysis data is also used in the calculation of 5-days backward air-mass trajectories with the HYSPLIT model using geopotential height, relative humidity, temperature, u/v wind components, vertical velocity at 37 pressure levels. The backward trajectories are initialized at 12 UTC, at each grid point of the study area and at a subregional mean cloud-top altitude obtained from the CALIPSO Level-2 5 km layer cloud product (version 3, daytime).

All meteorological variables are resampled from 0.5 degrees to the MODIS L3 resolution of 1 degrees and temporally averaged to the corresponding 8-day product. The temporal resolution of 8 days allows to combine large-scale~~and thermodynamic~~ , thermodynamic and aerosol forcings of cloud properties ~~, as~~ simultaneously on a synoptical scale. However, it must be taken into account that clouds adjust on different time scales (hours to several days) to ~~the large-scale circulation and their thermodynamic~~ their environment (Klein, 1997; McCoy et al., 2017; Adebiyi and Zuidema, 2018) and thus processes relevant on shorter time scales might be underrepresented in the data set.

2.2 Subregional GBRT models

In this study CF and REF are simulated based on a selected predictor set (AOD and meteorological parameters) in the SEA (10°–20° S, 0°–10° E, as analyzed in Klein and Hartmann (1993)) using Gradient Boosting Regression Trees (GBRTs). To account for subregional spatial variability of e.g. cloud altitude, aerosol occurrence, boundary layer dynamics and large-scale dynamics, the study area is divided into four equal-sized subregions of 5° by 5°: the northwestern (NW), northeastern (NE), southwestern (SW), southeastern (SE) subregion. Consequently, drivers of CF and REF are analyzed in the environmental context of each subregion individually, yielding eight ~~subregional statistical models~~ (four subregions x two predictands) subregional statistical models each based on approximately 2000 data points per parameter.

GBRTs are a highly robust machine learning technique aimed at mapping the relationship between a set of predictors and a predictand. The GBRT algorithm produces an ensemble of many weak prediction models (“base learners” or trees), which are expanded in stages, following the gradient descent of a specified loss function (Friedman, 2001; Natekin and Knoll, 2013). In each stage, a new decision tree is fitted to the residuals of the previous tree, and the prediction function is updated. The sum over all decision trees results in a robust statistical model that can map non-linear dependencies between predictors and the predictand. These statistical models are widely used in environmental and atmospheric sciences (e.g., Sayegh et al., 2016; Carslaw and Taylor, 2009) due to their predictive power, simple implementation and flexibility toward qualitative and quantitative data (Hastie et al., 2009). However, GBRTs require careful parameter tuning (e.g. boosting iterations, learning rate), as the goal is to represent the given data and relationships as accurately as possible, without overfitting the model. The GBRT implementation of the scikit-learn library was used and adapted to this end (Pedregosa et al., 2011).

To train, test and validate the statistical models, the data set is split into three random parts, the training (50 %), test (20 %) and validation (30 %) data sets. The model setup is tuned based on the training data by testing various scenarios specified by a parameter grid through 3-fold cross-validated search. During cross-validation, the training set is divided into three parts: two thirds are used for training and one third for testing. Each parameter combination from the grids, listed in Table 1, is evaluated

based on the r^2-score obtained in correlating predicted and observed output. The obtained hyper-parameter with the highest performance is chosen to set up the model. In general, a high number of boosting iterations and a low learning rate will increase the model's ability to ~~generalize~~make predictions on an unseen dataset (generalize), its performance and computational demand during training. The Huber loss function is chosen due to its higher robustness compared to other continuous loss functions,

Table 1. Model parameter grid tested during 3-fold cross validation.

model parameter	impact on model performance	parameter grid tested
learning rate	low values allow for better generalization	[0.05,0.04,0.03,0.02,0.01,0.009,0.007,0.003]
boosting iterations	large values improve performance, but risk overfitting	[2000,2400,2800,3200,3600,4000]
maximum depth of a tree	small numbers prevent overfitting	[2,3,4]
minimum samples per leaf	small values risk overfitting	[10,14,18]

e.g. least squares (Huber, 1964; Natekin and Knoll, 2013). A subsample rate (a random fraction of the training data used for fitting) of 0.8 is selected to reduce variance and increase model robustness. All remaining parameter settings are left at their default values as provided by the gradient boosting regressor function (Pedregosa et al., 2011).

Providing the optimal model setup, the model is fitted to the training data. In parallel, the test data set is used to regularize the GBRTs by determining the final boosting iteration. The learning stops when the mean squared error (MSE) of the test data set is increasing or constant five times in a row. The cross-validated tuning of the hyperparameter, the choice of a robust loss function and the implementation of an early stopping rule ensure the computing of robust GBRT models, which do not overfit to the training data.

To evaluate the overall performance of the GBRTs, two measures, the coefficient of determination (r^2) and the root mean squared error (RMSE) between predicted and observed CF (REF) are calculated using the independent validation dataset. To ensure comparability between the RMSE of the CF and REF performance the RMSE is normalized (NRMSE) by the difference between the maximum and minimum observed values.

The final model can be interpreted using 'partial dependence', which expresses the averaged change of a cloud property relative to a selected predictor set by averaging over all complement predictors (Friedman, 2001). This is done by computing an average prediction function for a given range of values (1^{st}–99^{th} percentile) estimated from the target predictor. Each grid point of the target predictor is fixed while the values of the complement predictors vary over their marginal probability density. As a result, the partial dependence represents the influence of one target variable, accounting for the full meteorological variation of the complement predictors. Accordingly, it is assumed that covarying cloud properties that are not explicitly considered in this study (e.g. liquid-water path) are indirectly constrained to some extent by the statistical model. This means that in the model the variation of meteorological parameters would implicitly represent different cloud states. The absolute difference of the maximum and minimum partial dependence is further used to compare the cloud property response due to the different predictors, thus to obtain a general measure for the most important drivers in the different subregions. In order to analyze the joined influence of two variables on the predictand, two-variable partial dependence plots are used. For regression trees the

implementation of partial dependence is straightforward and can be derived from the tree structure itself through a weighted tree traversal proposed by Friedman (2001). The partial dependence obtained from the GBRT model is added by the cloud property mean value for reference. 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.

In general, GBRTs are a powerful tool for representing non-linear dependencies and emphasize subregionally important determinants for low clouds in the SEA. However, for the interpretation it must be considered that partial dependencies rely on a statistical model. That means that associations between predictors and predictand are not necessarily causal, as in every statistical model. The obtained relationships are assumed to reflect mainly processes at a subspatial scale during the biomass-burning season, but may be to a small extent attributed to spatial and intraseasonal variations.

2.3 Predictor selection

The predictor selection pursues the goal of creating a simple model capable of capturing general thermodynamic, dynamic, stratification and aerosol patterns relevant for changes of cloud properties and is based on findings of previous studies (e.g., Norris and Iacobellis, 2005; Lacagnina and Selten, 2013; McCoy et al., 2017; Andersen et al., 2017; Fuchs et al., 2017; Adebiyi and Zuidema, 2018). Twelve predictors (see Table 2 for an overview) are chosen as inputs to the GBRTs due to their known forcing on CF and REF in the SEA. The listed parameters describe cloud-relevant environmental conditions at the sea surface (e.g., SST, MSLP), cloud level (RH950, RH850) and the free troposphere (e.g., T700, RH700).

Table 2. Predictors and abbreviations used in the GBRT models.

Thermodynamic properties	**Dynamic properties**
Lower tropospheric stability (LTS)	Source latitude of air mass (Lat_src)
Sea surface temperature (SST)	Source longitude of air mass (Lon_src)
Temperature at 700 hPa (T700)	Wind speed at 10m (WSP)
Relative humidity at 700 hPa (RH700),	Zonal wind speeds at 600 hPa (U600)
at 850 hPa (RH850), at 950 hPa (RH950)	Mean sea level pressure (MSLP)
Aerosol property	
Aerosol optical depth (AOD)	

The lower tropospheric stability, a proxy for inversion strength, and sea surface temperature are primary controls for the multi-day and seasonal cloud occurrence in the SEA (Klein and Hartmann, 1993; Klein et al., 1995; de Szoeke et al., 2016). Here, LTS is defined as the difference between potential temperature (θ) at 850 hPa and 1000 hPa as described in Painemal and Zuidema (2010).

As ~~relative humidity is essential for cloud formation processes and characteristics in the free troposphere and at cloud level~~ 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), ~~free tropospheric humidity~~ relative humidity values at 700~~hPa and at two different cloud levels (~~, 850 and 950 ~~and 850 hPa)~~ hPa are selected as ~~predictor~~predictors.

The large-scale circulation and the history of air masses drive boundary layer cloudiness (Klein et al., 1995; Mauger and Norris, 2007; Fuchs et al., 2017). In order to represent the influence of external dynamics on the local cloud field, the latitude and longitude of the origin of 5-day backward trajectories (Lat_src, Lon_src) are included as predictors in the statistical models. The backward trajectories are initiated at the mean cloud-top altitude in every subregion: 1090 m (NW), 1060 m (NE), 1180 m (SW), and 810 m (SE). Air-mass dynamics, including the surface wind speed and the strength of subtropical anticyclones, are important drivers for cloud amount, physical and radiative properties (Klein et al., 1995; Brueck et al., 2015; Kazil et al., 2015; Bretherton et al., 2013) and considered as predictors in the GBRT models. The strength of the South African Easterly Jet is observed to influence the marine boundary layer during the month of September to October through changes in stability and subsidence. It is defined as easterly wind speeds exceeding 6 m s^{-1} between 5° S to 15° S at 650–600 hPa (Adebiyi and Zuidema, 2016). In this study its influence is assumed to extend over the study area and thus the zonal wind field at 600 hPa is used.

Aerosols interact with liquid clouds in a multifaceted way (Fan et al., 2016). According to Twomey's theory of the first aerosol indirect effect, aerosols act as cloud condensation nuclei and influence cloud microphysics and albedo (Twomey, 1974). The Albrecht hypothesis states that this effect may result in a prolonged cloud lifetime, increased cloud optical thickness, liquid water path and cloud fraction through the suppression of precipitation (Albrecht, 1989). For the investigation of cloud susceptibility to aerosols, the AOD is considered as a proxy for cloud condensation nuclei. While the aerosol index may be a better proxy for cloud condensation nuclei than AOD (Stier, 2016), its computation requires the Ångström exponent, which is not available in the 8-day MODIS L3 product (Levy et al., 2013). Studies that observed the bivariate relations between AOD and cloud properties are numerous (e.g., Kaufman, 2005, 2006; Grandey et al., 2013), but spurious correlations exist. The strength of the relation between AOD and CF or REF is depending on satellite artifacts in the vicinity of clouds, e.g. cloud contamination and three-dimensional radiative effects (Várnai et al., 2013; Christensen et al., 2017) as well as on meteorological conditions, e.g. aerosol hygroscopic swelling with humidity (Kaufman et al., 2005; Quaas et al., 2010). In turn aerosols may alter the cloud's thermodynamic environment, through the semi-direct effect, where absorbing aerosol layers increase stability through local heating (Johnson et al., 2004; Li et al., 2013).

The ~~choice of predictors reflects the compromise between characterizing the atmospheric state sufficiently to derive~~ 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~~, but lacks interpretability. The application of the GBRTs aims at finding important meteorological controls for changes in CF and REF in the subregions of the SEA.~~ . 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.

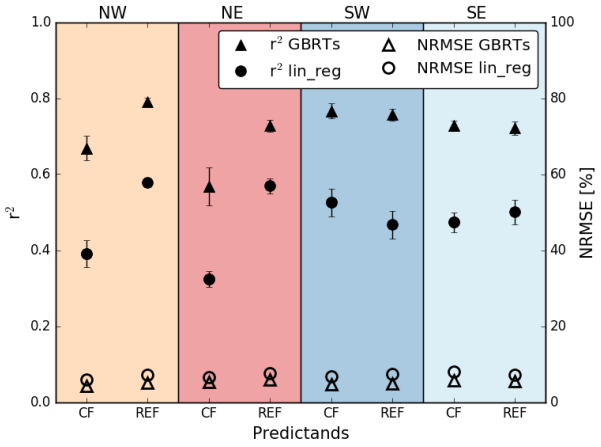

Figure 1. The overall mean quality of the GBRT models (triangles) is compared to a simple least squares linear regression (circles) for CF and REF in the four subregions NW, NE, SW and SE during JAS. The models are evaluated based on the coefficient of determination (r^2) and normalized root mean squared error (NRMSE) between predicted and observed CF (REF). The error bars range from the minimum to maximum r^2 obtained from 10 different models using randomly chosen training data.

3 Results and discussion

3.1 Validating GBRT models

In this section the statistical models are evaluated, important features within the models are identified and subsequently, partial dependencies (see Sect. 2.2 for more information) of the most important determinants are presented.

Figure 1 shows the validation results for the GBRTs predicting CF and REF in the different subregions. The performance is compared to a multiple linear regression analysis, using the same data basis. The correlation (r^2) of predicted and observed values in the GBRT model ranges from 0.57 to 0.79 in the different subregional models and is clearly superior to the r^2 of the multiple linear regression model ranging from 0.32 to 0.58 in the different subregions. The r^2 range (error bars) of 10 random GBRT simulations based on 10 different data random splits typically does not exceed the r^2 range of the linear regression using the 10 different data random splits, indicating constant model performances. Both models show a low normalized RMSE (NRMSE), that is on average $\sim$5 % for the GBRTs and $\sim$7 % for the linear regression.

Considering the GBRT models only, two aspects can be noted. First, in the northern subregions the REF models perform slightly better than the CF models, and second, the CF model shows subregional variations. Differences of model skills might be attributed to a higher variability of the cloud properties and meteorological conditions prevailing in the SW compared to the NE (Fuchs et al., 2017; Adebiyi and Zuidema, 2018; Rahn and Garreaud, 2010), or point to missing information in the predictor set of the NE-CF model.

As all GBRT models have been shown to adequately represent parameter relations, the statistical relationships within the models are subsequently analyzed with the purpose of inferring process relationships.

3.2 Sensitivity of cloud fraction and droplet radius

Figure 2 shows the multi-model mean absolute difference of the maximum and minimum partial dependence of CF (a) and REF (b) on the predictors as a measure for the sensitivity of these cloud properties to the various predictors.

In general, LTS, surface wind speed and relative humidity at 950 hPa play an important role for the determination of CF, however, marked subregional differences in their sensitivities can be identified (Fig. 2a). It is notable that ~~LTS~~ CF is most sensitive to ~~CF~~ LTS in the southern subregions. In the northeast, the impact of relative humidity at 950 hPa on CF is markedly reduced. Here, surface wind speed seems to be a key driver of CF. Changes in AOD seem to have a marked impact on CF only in the eastern subregions that are frequently exposed to high aerosol loadings.

The REF (Fig. 2b) is largely controlled by the free tropospheric temperature in the NE subregion. Here, REF is, similar to CF, strongly influenced by surface winds. In the SE, relative humidity at 950 hPa is an important driver for REF, while in the other subregions, relative humidity at 850 hPa has a stronger impact on REF due to the higher cloud level. In the SE, which is regularly exposed to the continuous warm and dry air advection from the coastal and continental region, an occasional moistening through dynamical changes may have a strong effect on cloud droplets of a thin cloud layer (Adebiyi et al., 2015).

The influence of dynamical parameters such as zonal wind at 600 hPa and air-mass origin (Lon_src) on REF is especially relevant in the SW, while LTS is a prominent influence in the NW. As expected, the contribution of aerosols to changes in CF and REF is small compared to the main meteorological drivers. However, the absolute differences indicate that aerosols appear to be most important for REF in the relatively pristine SW.

Based on these outcomes important predictors are brought into focus and the GBRT partial dependencies of CF and REF on selected predictor variables are analyzed in more detail in the following subsections.

3.2.1 Thermodynamics

In accordance with findings of earlier studies (Klein and Hartmann, 1993; Zhang et al., 2009), Fig. 3(a) shows that CF increases with LTS in all subregions. This relation is explained by reduced dry-air entrainment under stable conditions building a shallow, well-mixed and humid cloud layer (cf. Wood and Bretherton, 2006; Wood, 2012; Myers and Norris, 2013). Under very stable conditions, above 30 K temperature difference, the sensitivity of CF to LTS seems to be saturated and further stabilization does not increase the cloudiness anymore. This relates well to findings by Zhang et al. (2009), who detected the strongest CF sensitivity at intermediate LTS. It is remarkable that CF sensitivity to LTS in southern subregions is about twice as strong as in the northern subregions. This observation might be attributed to cloud break-up linked to midlatitude cyclones (Toniazzo et al., 2011; Fuchs et al., 2017). In contrast, in the NE the impact of LTS on CF is relatively weak as this area is characterized by more stable conditions with less thermodynamic variability.

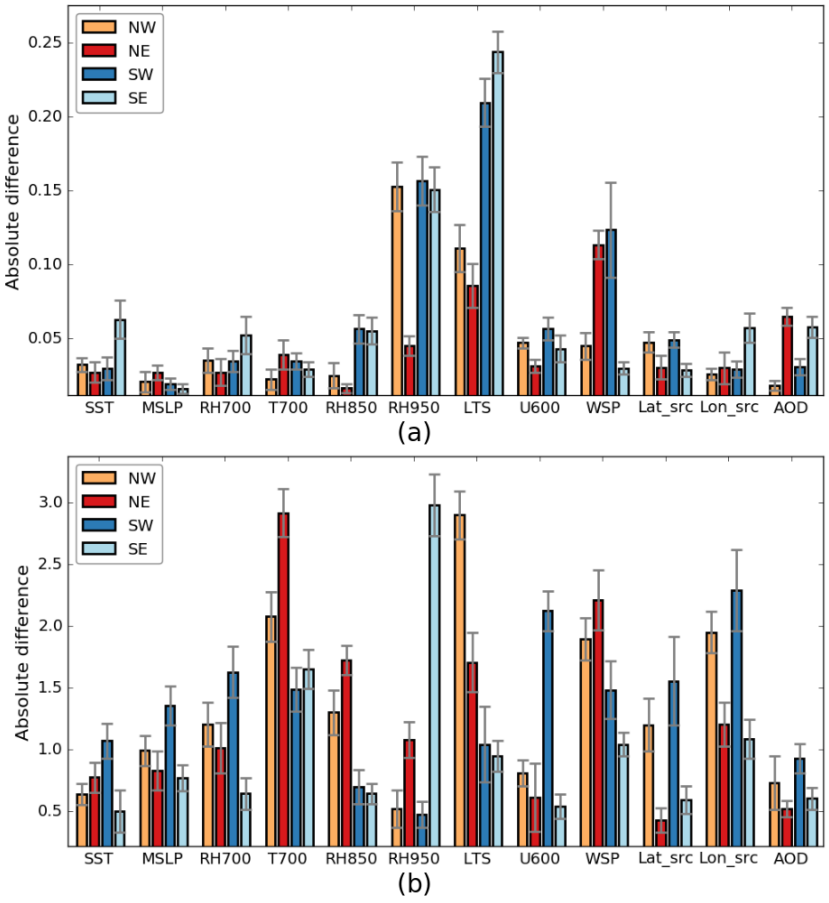

Figure 2. Mean absolute difference of maximum and minimum partial dependence of CF (a) and REF (b) on the predictors in the four subregions (colors) during JAS. 'Error' bars show the minimum and maximum absolute difference of partial dependencies of all model runs.

The relation of REF and LTS (Fig. 3b) is the strongest in the NW. A marked jump at ∼30 K may indicate the transition from a stable, relatively well-mixed coupled stratocumulus regime with larger droplets to an unstable, decoupled regime, where cloud liquid water evaporation due to dry and warm air entrainment can reduce droplet size (Bretherton and Wyant, 1997).

While the partial dependence of T700 on CF shows no distinct pattern in any subregion, a strong REF sensitivity to T700 can be noticed, in particular in the NE. As droplet size is retrieved at the cloud top, it might be more sensitive to a free tropospheric warming at 700 hPa and reduced dry-air entrainment above. The cloud cover, through the cloud's vertical extent, is probably more sensitive to the 850 hPa temperature, which is part of the LTS calculation (cf. Sect. 2.3).

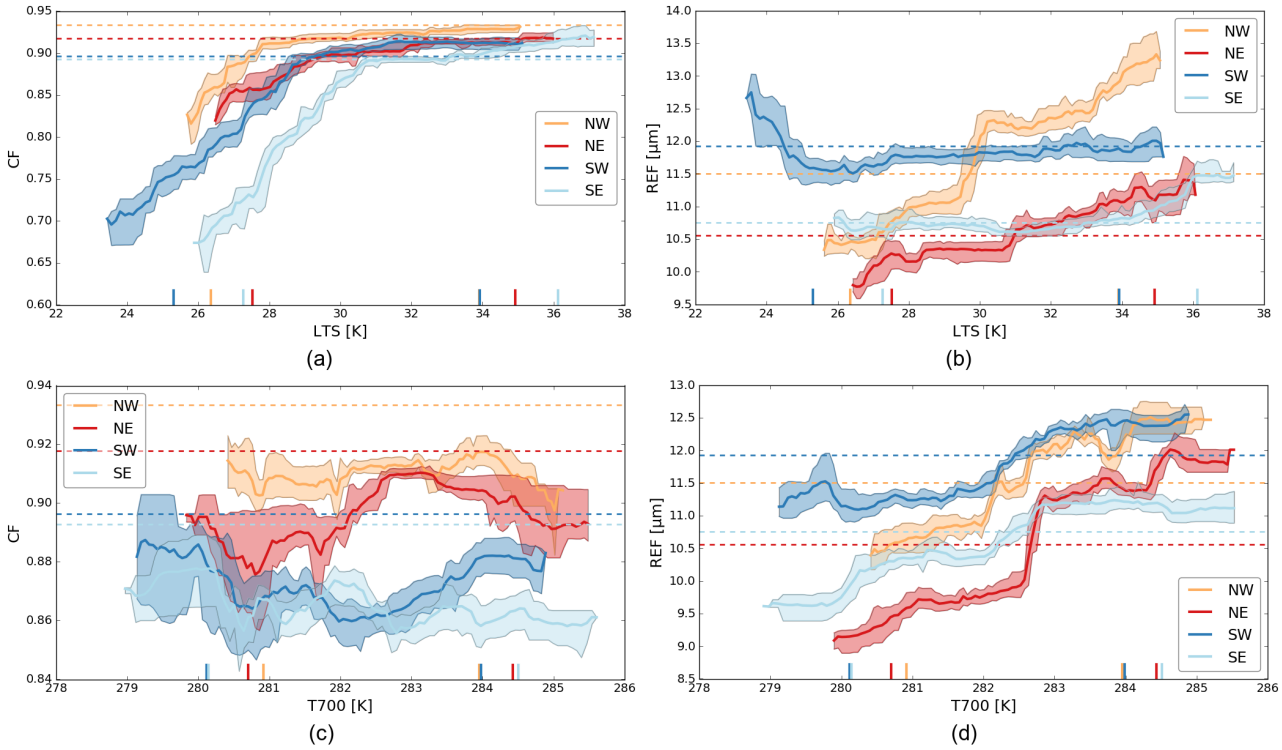

Figure 3. Mean partial dependence of CF and REF on LTS (a,b) and T700 (c,d) in the four subregions (colors) during JAS. Shaded areas mark minimum and maximum partial dependence obtained from all model runs. Horizontal dashed lines show the predicted mean. Vertical tick marks on the x-axis indicate $5th$ and $95th$ percentile of the observations.

3.2.2 Dynamics

Large-scale dynamics, here the origins of air masses, can influence cloud cover in the SEA in different ways (cf. Fuchs et al., 2017). Figure 4(a) shows the response of CF to changes in the latitudinal origin of air masses (Lat_src). While in the eastern subregions, CF seems largely insensitive to changes in Lat_src, CF in the western subregions is negatively associated with Lat_src: i.e. CF decreases the further north the air-mass origin. This likely points to findings of Fuchs et al. (2017), who found that long-distance air masses, induced by westerly disturbances, are related to increased boundary layer height, cloud fraction~~and~~ , cloud droplet sizes and liquid-water path in the western parts of the SEA. Air masses originating from ∼20° S (SW) and ∼15° S (NW) may contribute to the reduction of CF by subsiding dry air (Myers and Norris, 2013). The shift of the CF minima between southern and northern subregions may be interpreted as time lag of these air-mass paths, reaching the southern subregions earlier. In parallel to CF, the REF sensitivity to the latitudinal air-mass origin is particularly strong in the western subregions of the study area, especially the SW (Fig. 4b). The subregional difference between western and eastern subregions is even stronger than for CF. The NE shows only a weak response of REF to the latitudinal component of the air-mass origin due to the influence of mainly continental air-mass origins (Fuchs et al., 2017) ranging much more on the longitudinal scale

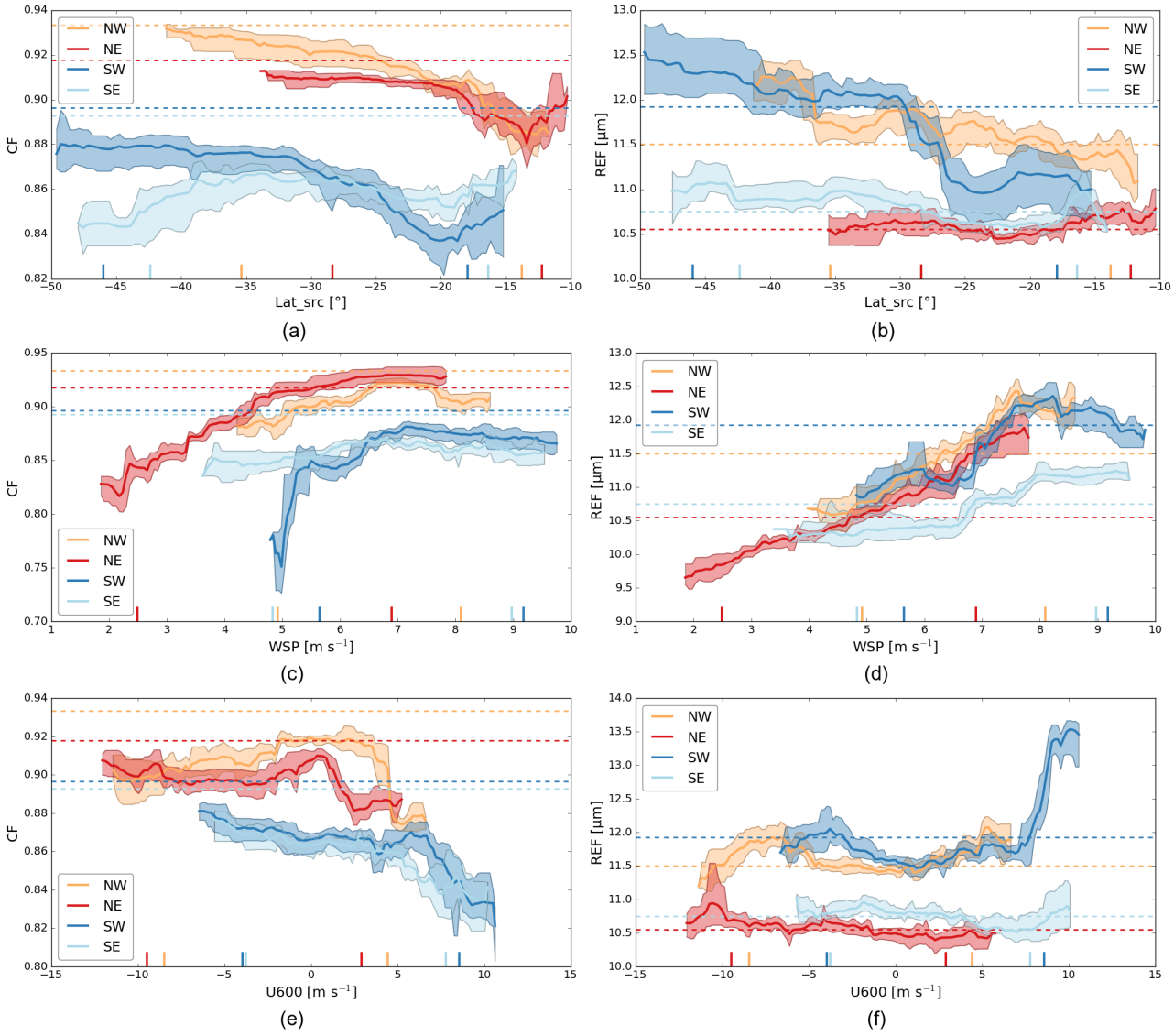

Figure 4. Mean partial dependence of CF and REF on source latitude of air mass (a,b), surface wind speed (c,d) and zonal winds at 600 hPa (e,f) in the four subregions (colors) during JAS. Details as in Fig. 3.

(Fig. 3b).

The CF sensitivity to the surface wind field is shown in Fig. 4(c). A clear increase of CF with higher surface wind speeds can be observed in the SW, where a change of wind speed of 1 m s^{-1} entails an increase of CF of more than 10 %. Strong surface winds may be associated with increased cold air advection and surface heat fluxes, favoring higher low-cloud amounts (Klein, 1997; Brueck et al., 2015). In all subregions, REF increases with wind speed (Fig. 4d), likely due to dynamic droplet growth in a more turbulent boundary layer. The partial dependence of CF on the zonal wind field at 600 hPa shows a decrease in

the southern subregions, when strong westerly winds are prevailing, and may indicate cloud-free areas in more convectively-driven systems (Fig. 4e). Weak tendencies of a CF enhancement in the southern subregions and a CF decrease in the NW due to stronger easterly winds are apparent and may indicate the influence of the South African Easterly Jet as discussed in Adebiyi and Zuidema (2016). As shown in Fig. 4(f), REF is largely insensitive towards the zonal wind fields at 600 hPa, presenting a strong effect only in the SW, where westerly winds are associated with larger droplets. These characteristics may support the effect of westerly disturbances which are more frequent in the SW.

Figure 5 shows the two-variable partial dependencies of REF on latitudinal and longitudinal air-mass origins for all four subregions, underlining regional differences in the susceptibility of REF to large-scale dynamical changes. In the SW, air masses originating from the far SW are connected to larger REF than air masses from the NE (Fig. 5c). In contrast, in the NE, larger REF are attributed to more humid air masses from the west (Fig. 5b), while easterly and probably drier winds from the continent favor smaller REF. The origin of air masses is more important for droplet size in the SW than in the NE through its higher subregional variability as a result of the occasional propagation of westerly disturbances.

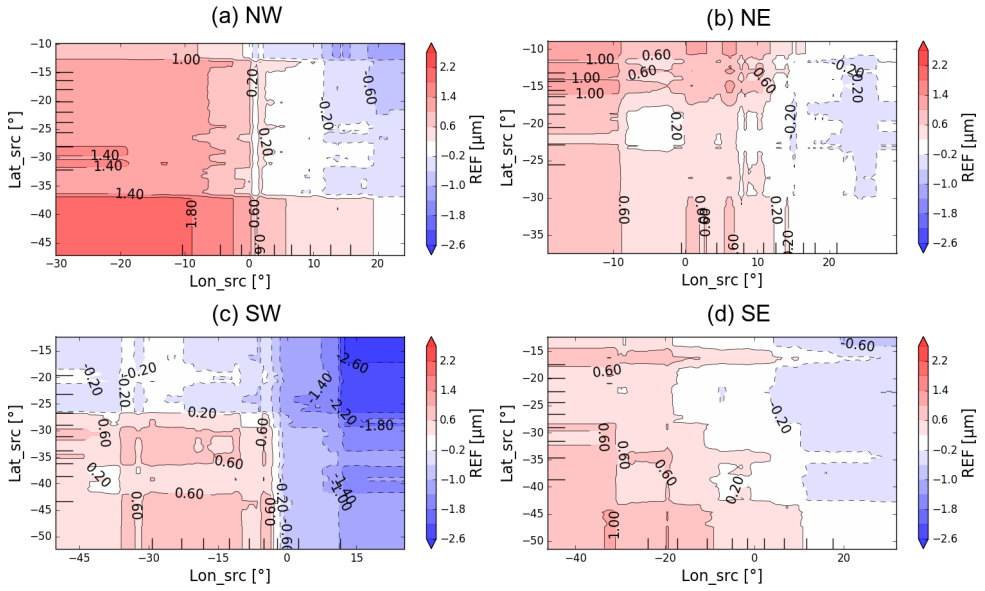

Figure 5. Two-variable partial dependence of REF on Lon_src and Lat_src in the in the four subregions ((a) NW, (b) NE, (c) SW, (d) SE) during JAS. Solid (dashed) contour lines indicate positive (negative) deviation of the predicted mean. The tick marks on the x-axis and y-axis indicate the deciles of the observations. For this illustration only one model run is selected at random ~~was used~~as it represents all model runs with error ranges comparable to that of the one-variable partial dependencies.

3.2.3 Conditions of aerosol-cloud interactions

Although the impact of aerosols on cloud properties ~~,~~ tends to be relatively weak on the temporal and spatial scales considered, characteristic patterns are obtained in the different subregions. CF increases with AOD in all subregions, especially in the southern subregions, as shown in Fig. 6(a). This relation is found in many studies and can have both ~~,~~ artificial and physical reasons (e.g., Mauger and Norris, 2007; Gryspeerdt et al., 2016; Andersen et al., 2017; Adebiyi and Zuidema, 2018). The observed relation may be physically induced through the availability of CCNs, increasing cloud lifetime and fractional cloudiness as aerosols are present (cf. Albrecht, 1989). It may be further explained by semi-direct effects, where absorbing carbonaceous aerosol layers heat the free troposphere causing a stabilization of the atmosphere that promotes the humidification of the cloud layer (cf. Li et al., 2013). Whether stability is enhanced by absorbing aerosols or is connected to the transport of aerosol-loaded warm air cannot be answered at this point. The effect of AOD enhancement on the AOD-CF relation due to hygroscopic swelling (Quaas et al., 2010) and wind-induced sea spray (Engström and Ekman, 2010), is thought to play a minor role due to the explicit consideration of relative humidity and surface wind speed in the statistical models. In the NE, the reason for the strong AOD-CF relation ($< 5th$ percentile of AOD) is intriguing but it is unclear to what extent it is caused by aerosol-related physical processes. It should be noted that these conditions only rarely occur.

The partial dependence of REF on the aerosol loading is shown in Fig. 6(b). The southern subregions show a comparable

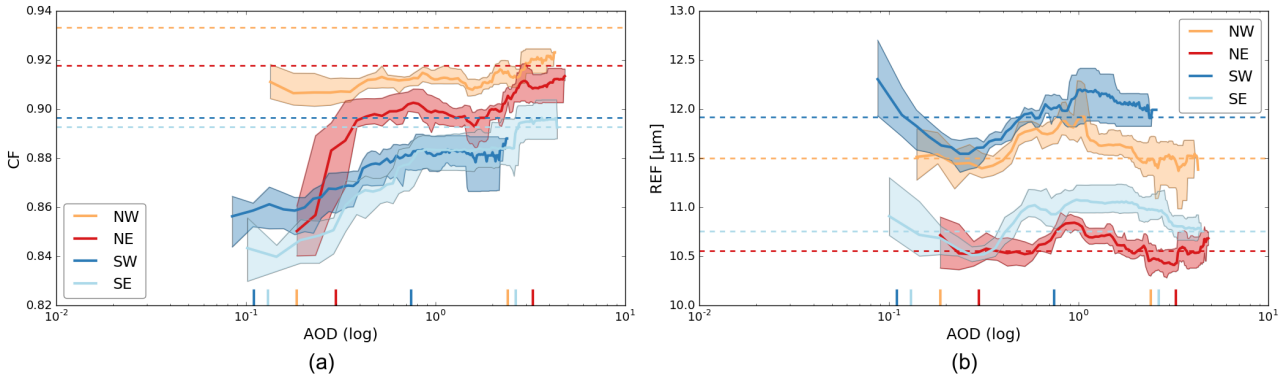

Figure 6. Mean partial dependence of CF (a) and REF (b) on AOD in the four subregions (colors) during JAS. Description as in Fig. 3.

pattern of a REF decrease up to AOD values of ∼0.2. A subsequent REF increase up to an AOD of ∼1 can be noticed in all subregions. The response of REF at lower AOD values is especially marked in the SW. Here, a different aerosol regime (composition and size: i.e. sea salt in the SW vs. biomass burning in the NE), giant cloud condensation nuclei, larger droplets in more turbulent conditions and the closer vicinity of aerosol and cloud layers may favor stronger aerosol indirect effects (cf. Andreae and Rosenfeld, 2008; Costantino and Bréon, 2012; Painemal et al., 2014; Andersen and Cermak, 2015). Stronger aerosol effects at low aerosol loadings were also found by Andersen et al. (2016) at a global scale. These results point to a saturation of the aerosol indirect effect under highly polluted conditions, where the influence of stability may be stronger. To what extent the relationship between REF and the AOD can be attributed to an absorbing aerosol bias in the satellite retrievals (Haywood et al., 2004) or physical processes cannot be answered definitively. However, the observed subregional differences of the polluted NE versus the more pristine SW make aerosol indirect effects more likely than retrieval issues.

Figure 7 shows AOD-REF partial dependencies for the months of July and September separately. While during July, REF seems to decrease with increasing AOD, especially in the SW subregion, during September the opposite relationship is found. The contrasting relationships may be related to differences in the vertical distribution of aerosols and clouds in the Southeast Atlantic. During July, aerosol and cloud layers are frequently entangled, facilitating ACI, whereas in September they can be well separated (Adebiyi et al. 2018). During this time, absorbing aerosol may increase the stability and trap humidity in the boundary layer, potentially leading to the observed relationship. The JAS partial dependence between AOD and REF can thus be viewed as a summary of these patterns. However, it is not the study's focus to separate the different aerosol effects mentioned earlier, but to analyze the overall influence of aerosols on clouds during the biomass-burning season.

The two-variable partial dependencies, presented in Fig. 8 to 12, show how the sensitivities of CF and REF to aerosol loading may vary under different meteorological conditions, i.e. LTS and relative humidity at 950 and 850 hPa. All subregions of the SEA are characterized by a stronger CF (Fig. 8) and REF sensitivity (Fig. 9) to LTS compared to AOD. In the southern subregions, CF is increased under stable and strongly polluted conditions. Here, the increase of CF with AOD is more pronounced in stable conditions, presumably due to reduced dry-air entrainment (cf. Chen et al., 2014), while CF seems to be less sensitive to

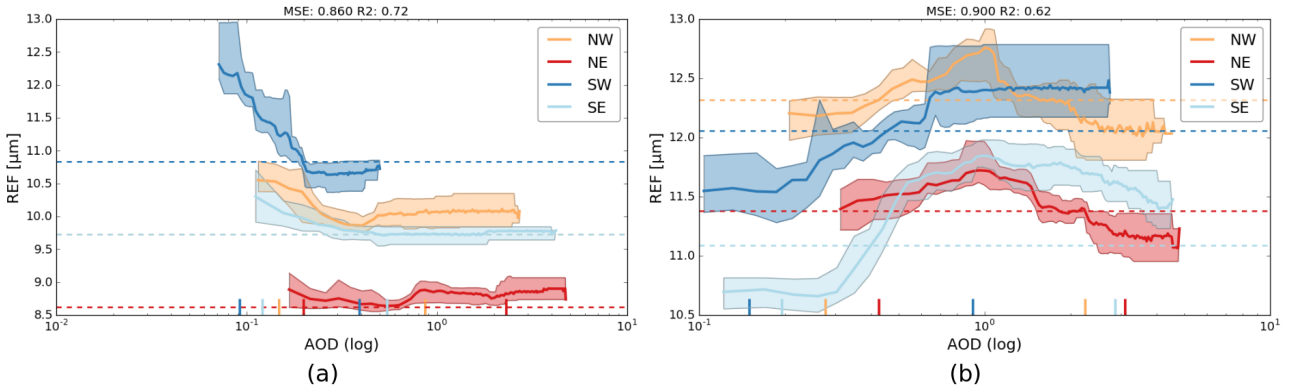

Figure 7. Mean partial dependence of REF on AOD in the four subregions (colors) in July (a) and September (b). MSE and r^2 refer to monthly GBRT models based on ∼600 data points. Description as in Fig. 3.

aerosols in unstable conditions, where primarily low CF may result from cloud breakups in midlatitude cyclones (cf. Toniazzo et al., 2011). In contrast, a generally higher REF sensitivity to aerosols characterizes the SW. In this subregion, larger droplets may more effectively persist, grow, and are thus susceptible to aerosols in both, stable and unstable (mixing of aerosols into the cloud layer) conditions (cf. Painemal et al., 2014). In the NE, it can further be observed that the CF sensitivity to aerosols is favored at low aerosol loading, which might be explained by the saturation of aerosol effects at higher loading (cf. de Szoeke et al., 2016).

The relation of CF (REF), humidity at 950 hPa and AOD is shown in Fig. 10 (11). Humidity at 950 hPa dominates all subregions, particularly the SE, while the impact of aerosols is relatively small. In the southern subregions, though, CF increases under humid and polluted conditions (Fig. 10c,d). CF is especially sensitive to an increase of aerosol loading below a cloud level-humidity of ∼80 %, while above this level aerosol swelling is more likely to affect the AOD retrieval (cf. Adebiyi and Zuidema, 2018). As shown for CF, relative humidity is essentially related to REF, and a reduction of REF due to aerosols is apparent throughout the different humidity ranges at 850 and 950 hPa (Fig. 11 and 12). In the SW (Fig. 12c), REF may be sensitive in drier as well as more humid conditions: while humid conditions provide larger droplets, entrainment induced by aerosols may more effectively reduce droplet size in dry conditions (cf. Chen et al., 2014).

In sum, the presented results show the potential of observing ACI susceptibilities in different thermodynamic conditions. Nevertheless, the presented link between meteorological conditions and aerosol effect on clouds (indirect and semi-direct) is not necessarily causal and further effects due to aerosol processing near clouds and satellite artifacts (Sect. 2.3) may contribute to the observed cloud sensitivities.

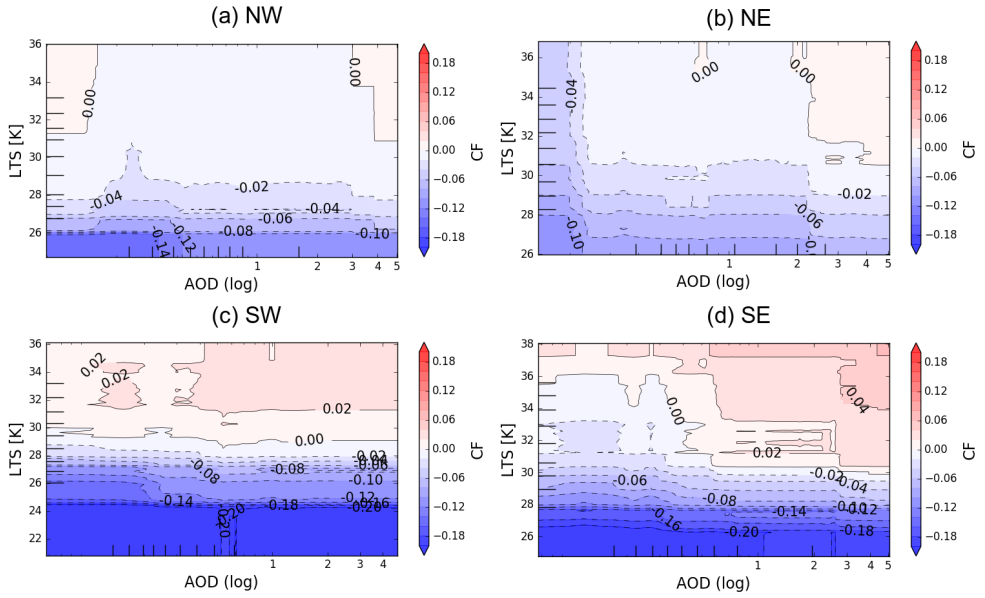

Figure 8. Two-variable partial dependence of CF on LTS and AOD in the four subregions NW (a), NE (b), SW (c), SE (d) during JAS. Description as in Fig. 5.

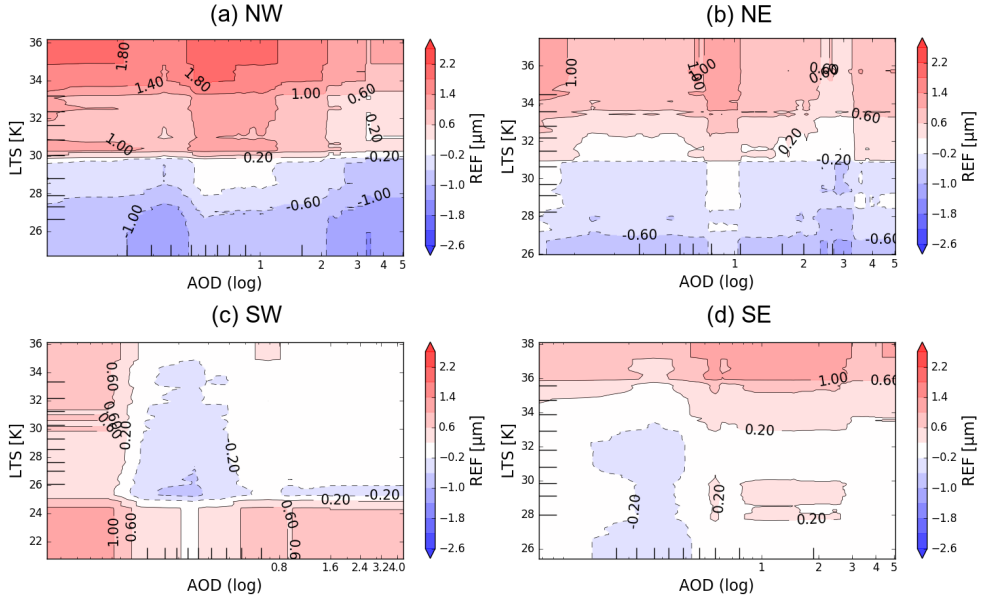

Figure 9. Two-variable partial dependence of REF on LTS and AOD in the four subregions NW (a), NE (b), SW (c), SE (d) during JAS. Description as in Fig. 5.

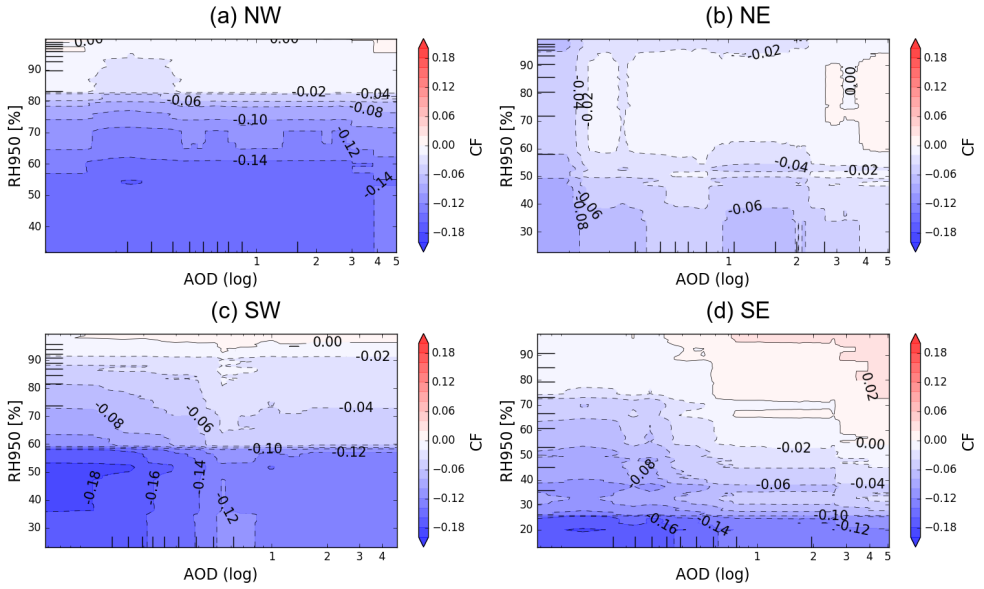

Figure 10. Two-variable partial dependence of CF on RH950 and AOD in the four subregions NW (a), NE (b), SW (c), SE (d) during JAS. Description as in Fig. 5.

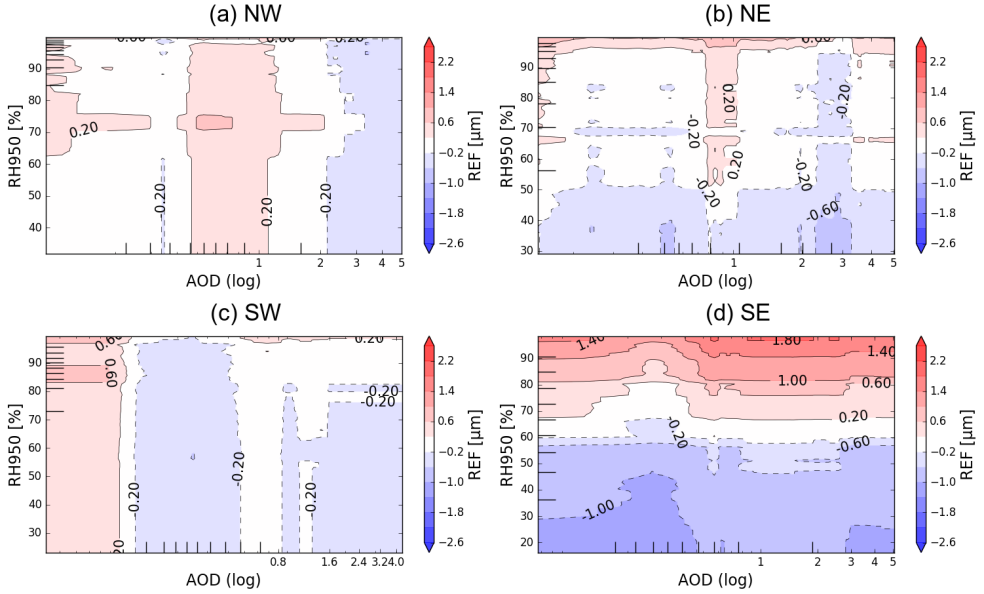

Figure 11. Two-variable partial dependence of REF on RH950 and AOD in the four subregions NW (a), NE (b), SW (c), SE (d) during JAS. Description as in Fig. 5.

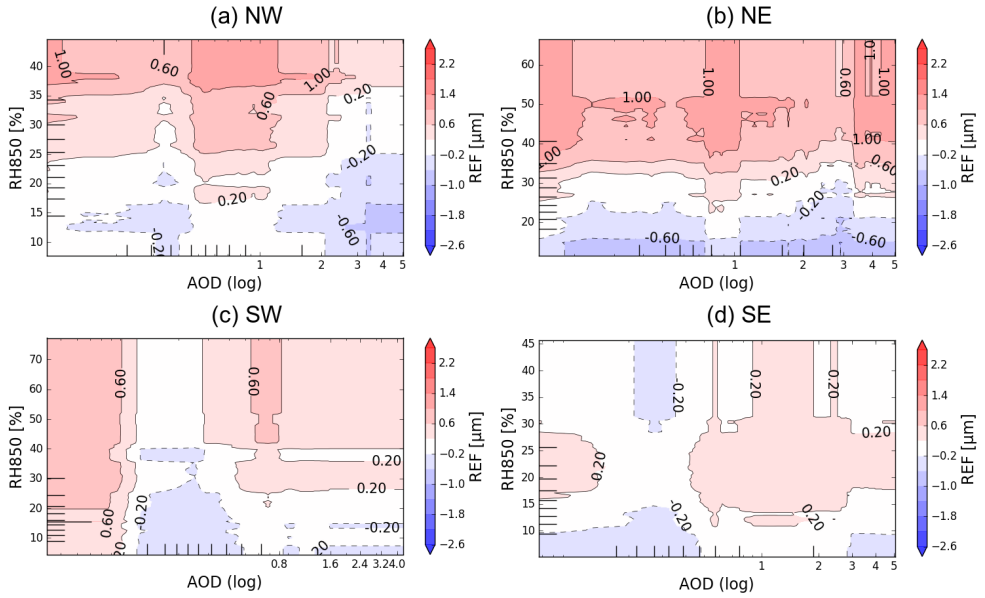

Figure 12. Two-variable partial dependence of REF on RH850 and AOD in the four subregions NW, NE, SW, SE during JAS. Description as in Fig. 5.

4 Conclusions

In this study relevant mechanisms for changes in CF and REF are analyzed by using a GBRT model in four subregions of the Southeast Atlantic. The GBRT models perform significantly better than multiple regression analyses based on the same data (average r^2 of 0.72 vs. 0.48, respectively). This indicates that the GBRT models can be used to adequately represent the interactions governing the cloud system, while the methodical approach proves advantageous. The model skill varies with subregion and cloud property and features different sensitivities to the same predictor set. Outcomes of the GBRTs provide useful insights of important determinants for cloud properties. By accounting for meteorological conditions and aerosol loadings the models can help untangling the various cloud processes and cloud sensitivities to aerosols in the subregions of the SEA. The subregional importance and patterns of cloud drivers and ACI sensitivities is plausible and in accordance with findings of related studies (e.g., Chen et al., 2014; McCoy et al., 2017; Adebiyi and Zuidema, 2018).

In the statistical models atmospheric stability, air-mass dynamics and relative humidity at cloud level are the most important drivers for changes in CF and REF, relative to the given set of predictors. The SEA cloud cover is dominated by LTS in all subregions. In the NE, cloud amount and droplet size is additionally controlled by surface wind speeds, while in the SE, both are essentially influenced by the availability of moisture. Large-scale dynamics is the main driver of changes of cloud properties in the SW.

The positive relation between LTS and CF obtained from the GBRT models is explained by the stabilization of the boundary layer dynamics, which promotes cloud amount and longevity. The sensitivity of CF to LTS is non-linear and saturates in stable

conditions of LTS > ∼30 K. LTS is especially important in the southern subregions, which are exposed to more variable atmospheric states.

Air-mass dynamics (air-mass origin and zonal wind speeds at 600 hPa) determine REF in the SW to a greater extent than in the NE. The REF increase in the SW is attributed to the outreach of convective westerly disturbances to this subregion. In the NE, air masses show less variability as they approach mainly from the continent under more stable conditions. Here, dynamically induced strong wind speeds and a warm free troposphere are associated with larger droplets.

Although aerosols play a secondary role for the prediction of cloud properties, important implications for the subregional strength of ACI can be derived from the model's partial dependencies. In the southern subregions, a strong sensitivity of CF and REF to AOD is modeled, likely due to aerosol-cloud interactions and semi-direct effects. CF sensitivities to aerosols are shown to be stronger in stable conditions, where dry-air entrainment is reduced. A higher REF sensitivity in unstable conditions is attributed to e.g. generally larger droplets, a different aerosol composition (e.g. sea salt) and a more turbulent layer, which possibly favors stronger aerosol indirect effects in these regions. Outcomes also point to the saturation of the aerosol indirect effect in the NE compared to the SW where low aerosol loadings may more efficiently act as cloud condensation nuclei.

This study presents the potential of using multivariate GBRTs to derive cloud determinants, and non-linear sensitivities and further to give realistic estimates of the magnitude of aerosol relationships on a synoptic scale. Due to the limited capability of a statistical model to learn the data inherent relations only, feedback mechanisms and satellite artifacts in the SEA cannot completely be accounted for. However, the application of machine learning techniques is advantageous and yields valuable insights into subregional cloud and ACI processes on the microphysical and macrophysical scale.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. MODIS data were obtained from the Goddard Space Flight Center (http://ladsweb.nascom.nasa.gov/data/search.html). The authors gratefully acknowledge the NOAA Air Resources Laboratory (ARL) for the provision of the HYSPLIT transport model (http://ready.arl.noaa.gov/HYSPLIT.php). ERA-Interim data were obtained from the homepage of European Centre for Medium-Range Weather Forecasts (http://apps.ecmwf.int). CALIPSO data were accessed through the NASA Langley Research Center Atmospheric Science Data Center (https://eosweb.larc.nasa.gov). The contribution of Hendrik Andersen was supported by Deutsche Forschungsgemeinschaft (DFG) in the project Namib Fog Life Cycle Analysis (NaFoLiCA), CE 163/7-1. The valuable comments of two anonymous reviewers helped improve the original manuscript.

References

Adebiyi, A. A. and Zuidema, P.: The role of the southern African easterly jet in modifying the southeast Atlantic aerosol and cloud environments, Quarterly Journal of the Royal Meteorological Society, 142, 1574–1589, https://doi.org/10.1002/qj.2765, http://doi.wiley.com/10.1002/qj.2765, 2016.

Adebiyi, A. A. and Zuidema, P.: Low Cloud Cover Sensitivity to Biomass-Burning Aerosols and Meteorology over the Southeast Atlantic, Journal of Climate, 31, 4329–4346, https://doi.org/10.1175/JCLI-D-17-0406.1, http://journals.ametsoc.org/doi/10.1175/JCLI-D-17-0406.1, 2018.

Adebiyi, A. A., Zuidema, P., and Abel, S. J.: The Convolution of Dynamics and Moisture with the Presence of Shortwave Absorbing Aerosols over the Southeast Atlantic, Journal of Climate, 28, 1997–2024, https://doi.org/10.1175/JCLI-D-14-00352.1, http://journals.ametsoc.org/doi/10.1175/JCLI-D-14-00352.1, 2015.

Albrecht, B. A.: Aerosols, Cloud Microphysics, and Fractional Cloudiness, Science, 245, 1227–1230, https://doi.org/10.1126/science.245.4923.1227, http://www.sciencemag.org/cgi/doi/10.1126/science.245.4923.1227, 1989.

Andersen, H. and Cermak, J.: How thermodynamic environments control stratocumulus microphysics and interactions with aerosols, Environmental Research Letters, 10, 024 004, https://doi.org/10.1088/1748-9326/10/2/024004, http://dx.doi.org/10.1088/1748-9326/10/2/024004http://stacks.iop.org/1748-9326/10/i=2/a=024004?key=crossref.a6bbe1459a1a562a56204f16af158cc2, 2015.

Andersen, H., Cermak, J., Fuchs, J., and Schwarz, K.: Global observations of cloud-sensitive aerosol loadings in low-level marine clouds, Journal of Geophysical Research: Atmospheres, 121, 12,936–12,946, https://doi.org/10.1002/2016JD025614, http://doi.wiley.com/10.1002/2016JD025614, 2016.

Andersen, H., Cermak, J., Fuchs, J., Knutti, R., and Lohmann, U.: Understanding the drivers of marine liquid-water cloud occurrence and properties with global observations using neural networks, Atmospheric Chemistry and Physics, 17, 9535–9546, https://doi.org/10.5194/acp-17-9535-2017, https://doi.org/10.5194/acp-17-9535-2017https://www.atmos-chem-phys.net/17/9535/2017/, 2017.

Andreae, M. and Rosenfeld, D.: Aerosol-cloud-precipitation interactions. Part 1. The nature and sources of cloud-active aerosols, Earth-Science Reviews, 89, 13–41, https://doi.org/10.1016/j.earscirev.2008.03.001, http://www.ncbi.nlm.nih.gov/pubmed/11713467http://www.nature.com/articles/nn766http://linkinghub.elsevier.com/retrieve/pii/S0012825208000317, 2008.

Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Berntsen, T., DeAngelo, B. J., Flanner, M. G., Ghan, S., Kärcher, B., Koch, D., Kinne, S., Kondo, Y., Quinn, P. K., Sarofim, M. C., Schultz, M. G., Schulz, M., Venkataraman, C., Zhang, H., Zhang, S., Bellouin, N., Guttikunda, S. K., Hopke, P. K., Jacobson, M. Z., Kaiser, J. W., Klimont, Z., Lohmann, U., Schwarz, J. P., Shindell, D., Storelvmo, T., Warren, S. G., and Zender, C. S.: Bounding the role of black carbon in the climate system: A scientific assessment, Journal of Geophysical Research: Atmospheres, 118, 5380–5552, https://doi.org/10.1002/jgrd.50171, http://doi.wiley.com/10.1002/jgrd.50171, 2013.

Bony, S. and Dufresne, J.-L.: Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models, Geophysical Research Letters, 32, L20 806, https://doi.org/10.1029/2005GL023851, http://doi.wiley.com/10.1029/2005GL023851, 2005.

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., Kerminen, V.-M., Kondo, Y., Liao, H., Lohmann, U., Rasch, P., Satheesh, S. K., Sherwood, S., Stevens, B., Zhang, X. Y., and Zhan, X. Y.: Clouds and Aerosols, in: Climate Change 2013 - The Physical Science Basis, edited by Intergovernmental Panel on Climate Change, chap. 7, pp. 571–658, Cambridge University Press, Cambridge, https://doi.org/10.1017/CBO9781107415324.016, http://ebooks.cambridge.org/ref/id/CBO9781107415324A024https://www.cambridge.org/core/product/identifier/CBO9781107415324A024/type/book{_}part, 2013.

Bretherton, C. S. and Wyant, M. C.: Moisture Transport, Lower-Tropospheric Stability, and Decoupling of Cloud-Topped Boundary Layers, Journal of the Atmospheric Sciences, 54, 148–167, https://doi.org/10.1175/1520-0469(1997)054<0148:MTLTSA>2.0.CO;2, http://journals.ametsoc.org/doi/abs/10.1175/1520-0469{%}281997{%}29054{%}3C0148{%}3AMTLTSA{%}3E2.0.CO{%}3B2, 1997.

Bretherton, C. S., Blossey, P. N., and Jones, C. R.: 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, 316–337, https://doi.org/10.1002/jame.20019, 2013.

Brueck, M., Nuijens, L., and Stevens, B.: On the Seasonal and Synoptic Time-Scale Variability of the North Atlantic Trade Wind Region and Its Low-Level Clouds, Journal of the Atmospheric Sciences, 72, 1428–1446, https://doi.org/10.1175/JAS-D-14-0054.1, http://journals.ametsoc.org/doi/10.1175/JAS-D-14-0054.1, 2015.

Carslaw, D. C. and Taylor, P. J.: Analysis of air pollution data at a mixed source location using boosted regression trees, Atmospheric Environment, 43, 3563–3570, https://doi.org/10.1016/j.atmosenv.2009.04.001, http://dx.doi.org/10.1016/j.atmosenv.2009.04.001, 2009.

Chand, D., Wood, R., Anderson, T. L., Satheesh, S. K., and Charlson, R. J.: Satellite-derived direct radiative effect of aerosols dependent on cloud cover, Nature Geoscience, 2, 181–184, https://doi.org/10.1038/ngeo437, http://www.nature.com/articles/ngeo437, 2009.

Chen, Y.-C., Christensen, M. W., Stephens, G. L., and Seinfeld, J. H.: Satellite-based estimate of global aerosol-cloud radiative forcing by marine warm clouds, Nature Geoscience, 7, 643–646, https://doi.org/10.1038/ngeo2214, http://www.nature.com/articles/ngeo2214, 2014.

Christensen, M. W., Neubauer, D., Poulsen, C. A., Thomas, G. E., McGarragh, G. R., Povey, A. C., Proud, S. R., and Grainger, R. G.: Unveiling aerosol–cloud interactions – Part 1: Cloud contamination in satellite products enhances the aerosol indirect forcing estimate, Atmospheric Chemistry and Physics, 17, 13 151–13 164, https://doi.org/10.5194/acp-17-13151-2017, https://www.atmos-chem-phys.net/17/13151/2017/, 2017.

Costantino, L. and Bréon, F.-M.: Aerosol indirect effect on warm clouds over South-East Atlantic, from co-located MODIS and CALIPSO observations, Atmospheric Chemistry and Physics Discussions, 12, 14 197–14 246, https://doi.org/10.5194/acpd-12-14197-2012, http://www.atmos-chem-phys-discuss.net/12/14197/2012/, 2012.

de Szoeke, S. P., Verlinden, K. L., Yuter, S. E., and Mechem, D. B.: The time scales of variability of marine low clouds, Journal of Climate, 29, 6463–6481, https://doi.org/10.1175/JCLI-D-15-0460.1, 2016.

Eastman, R., Wood, R., and Bretherton, C. S.: Time Scales of Clouds and Cloud-Controlling Variables in Subtropical Stratocumulus from a Lagrangian Perspective, Journal of the Atmospheric Sciences, 73, 3079–3091, https://doi.org/10.1175/JAS-D-16-0050.1, http://journals.ametsoc.org/doi/10.1175/JAS-D-16-0050.1, 2016.

Engström, A. and Ekman, A. M. L.: Impact of meteorological factors on the correlation between aerosol optical depth and cloud fraction, Geophysical Research Letters, 37, https://doi.org/10.1029/2010GL044361, http://doi.wiley.com/10.1029/2010GL044361, 2010.

Fan, J., Wang, Y., Rosenfeld, D., Liu, X., Fan, J., Wang, Y., Rosenfeld, D., and Liu, X.: Review of Aerosol-Cloud Interactions: Mechanisms, Significance, and Challenges, Journal of the Atmospheric Sciences, 73, 4221–4252, https://doi.org/10.1175/JAS-D-16-0037.1, http://journals.ametsoc.org/doi/10.1175/JAS-D-16-0037.1, 2016.

Friedman, J. H.: Greedy function approximation: A gradient boosting machine, Annals of Statistics, 29, 1189–1232, https://doi.org/DOI 10.1214/aos/1013203451, 2001.

Fuchs, J., Cermak, J., Andersen, H., Hollmann, R., and Schwarz, K.: On the Influence of Air Mass Origin on Low-Cloud Properties in the Southeast Atlantic, Journal of Geophysical Research: Atmospheres, 122, 11,076–11,091, https://doi.org/10.1002/2017JD027184, https://doi.org/10.1002/2017JD027184http://doi.wiley.com/10.1002/2017JD027184, 2017.

Grandey, B. S., Stier, P., and Wagner, T. M.: Investigating relationships between aerosol optical depth and cloud fraction using satellite, aerosol reanalysis and general circulation model data, Atmospheric Chemistry and Physics, 13, 3177–3184, https://doi.org/10.5194/acp-13-3177-2013, http://www.atmos-chem-phys.net/13/3177/2013/, 2013.

Gryspeerdt, E., Quaas, J., and Bellouin, N.: Constraining the aerosol influence on cloud fraction, Journal of Geophysical Research: Atmospheres, 121, 3566–3583, https://doi.org/10.1002/2015JD023744, http://doi.wiley.com/10.1002/2015JD023744, 2016.

Hastie, T., Tibshirani, R., and Friedman, J.: The Elements of Statistical Learning, Springer Series in Statistics, Springer New York, New York, NY, 2 edn., https://doi.org/10.1007/978-0-387-84858-7, https://web.stanford.edu/{~}hastie/Papers/ESLII.pdfhttp://www.springerlink.com/index/D7X7KX6772HQ2135.pdf{%}255Cnhttp://www-stat.stanford.edu/{~}tibs/book/preface.pshttp://www.tandfonline.com/doi/abs/10.1198/jasa.2004.s339http://link.springer.com/10.1007/978, 2009.

Haywood, J. M., Osborne, S. R., and Abel, S. J.: The effect of overlying absorbing aerosol layers on remote sensing retrievals of cloud effective radius and cloud optical depth, Quarterly Journal of the Royal Meteorological Society, 130, 779–800, https://doi.org/10.1256/qj.03.100, http://doi.wiley.com/10.1256/qj.03.100, 2004.

Hubanks, P., King, M., Platnick, S., and Pincus, R.: MODIS Atmosphere L3 Gridded Product Algorithm Theoretical Basis Document No. ATBD-MOD-30 for Level-3 Global Gridded Atmosphere Products (08_D3, 08_E3, 08_M3) and Users Guide, http://modis-atmos.gsfc.nasa.gov/MOD08{_}M3/atbd.html, 2018.

Huber, P. J.: Robust Estimation of a Location Parameter, The Annals of Mathematical Statistics, 35, 73–101, https://doi.org/10.1214/aoms/1177703732, http://projecteuclid.org/euclid.aoms/1177703732, 1964.

Johnson, B. T., Shine, K. P., and Forster, P. M.: The semi-direct aerosol effect: Impact of absorbing aerosols on marine stratocumulus, Quarterly Journal of the Royal Meteorological Society, 130, 1407–1422, https://doi.org/10.1256/qj.03.61, http://doi.wiley.com/10.1256/qj.03.61, 2004.

Jones, C. R., Bretherton, C. S., and Blossey, P. N.: Fast stratocumulus time scale in mixed layer model and large eddy simulation, Journal of Advances in Modeling Earth Systems, pp. 206–222, https://doi.org/10.1002/2013MS000289.Received, http://onlinelibrary.wiley.com/doi/10.1002/2013MS000289/full, 2014.

Kaufman, Y., Remer, L., Tanre, D., Rong-Rong Li, Kleidman, R., Mattoo, S., Levy, R., Eck, T., Holben, B., Ichoku, C., Martins, J., and Koren, I.: A critical examination of the residual cloud contamination and diurnal sampling effects on MODIS estimates of aerosol over ocean, IEEE Transactions on Geoscience and Remote Sensing, 43, 2886–2897, https://doi.org/10.1109/TGRS.2005.858430, http://ieeexplore.ieee.org/document/1542360/, 2005.

Kaufman, Y. J.: Dust transport and deposition observed from the Terra-Moderate Resolution Imaging Spectroradiometer (MODIS) spacecraft over the Atlantic Ocean, Journal of Geophysical Research, 110, D10S12, https://doi.org/10.1029/2003JD004436, http://www.agu.org/pubs/crossref/2005/2003JD004436.shtmlhttp://doi.wiley.com/10.1029/2003JD004436, 2005.

Kaufman, Y. J.: Smoke and Pollution Aerosol Effect on Cloud Cover, Science, 313, 655–658, https://doi.org/10.1126/science.1126232, http://www.sciencemag.org/cgi/doi/10.1126/science.1126232, 2006.

Kazil, J., Feingold, G., and Yamaguchi, T.: Wind speed response of marine non-precipitating stratocumulus clouds over a diurnal cycle in cloud-system resolving simulations, Atmospheric Chemistry and Physics Discussions, 15, 28 395–28 452, https://doi.org/10.5194/acpd-15-28395-2015, http://www.atmos-chem-phys-discuss.net/15/28395/2015/, 2015.

Klein, S. and Hartmann, D.: The seasonal cycle of low stratiform clouds, Journal of Climate, 6, 1587–1606, http://journals.ametsoc.org/doi/abs/10.1175/1520-0442(1993)006{%}3C1587:TSCOLS{%}3E2.0.CO;2, 1993.

Klein, S. A.: Synoptic Variability of Low-Cloud Properties and Meteorological Parameters in the Subtropical Trade Wind Boundary Layer, Journal of Climate, 10, 2018–2039, https://doi.org/10.1175/1520-0442(1997)010<2018:SVOLCP>2.0.CO;2, https://www.gfdl.noaa.gov/bibliography/related{_}files/sak9701.pdfhttp://journals.ametsoc.org/doi/abs/10.1175/1520-0442{%}281997{%}29010{%}3C2018{%}3ASVOLCP{%}3E2.0.CO{%}3B2, 1997.

Klein, S. A., Hartmann, D. L., and Norris, J. R.: On the Relationships among Low-Cloud Structure, Sea Surface Temperature, and Atmospheric Circulation in the Summertime Northeast Pacific, Journal of Climate, 8, 1140–1155, https://doi.org/10.1175/1520-0442(1995)008<1140:OTRALC>2.0.CO;2, http://journals.ametsoc.org/doi/abs/10.1175/1520-0442{%}281995{%}29008{%}3C1140{%}3AOTRALC{%}3E2.0.CO{%}3B2, 1995.

Lacagnina, C. and Selten, F.: A novel diagnostic technique to investigate cloud-controlling factors, Journal of Geophysical Research: Atmospheres, 118, 5979–5991, https://doi.org/10.1002/jgrd.50511, http://doi.wiley.com/10.1002/jgrd.50511, 2013.

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., and Hsu, N. C.: The Collection 6 MODIS aerosol products over land and ocean, Atmospheric Measurement Techniques, 6, 2989–3034, https://doi.org/10.5194/amt-6-2989-2013, http://www.atmos-meas-tech.net/6/2989/2013/, 2013.

Li, J., Von Salzen, K., Peng, Y., Zhang, H., and Liang, X. Z.: Evaluation of black carbon semi-direct radiative effect in a climate model, Journal of Geophysical Research Atmospheres, 118, 4715–4728, https://doi.org/10.1002/jgrd.50327, 2013.

Mauger, G. S. and Norris, J. R.: Meteorological bias in satellite estimates of aerosol-cloud relationships, Geophysical Research Letters, 34, https://doi.org/10.1029/2007GL029952, http://doi.wiley.com/10.1029/2007GL029952, 2007.

Mauger, G. S. and Norris, J. R.: Assessing the Impact of Meteorological History on Subtropical Cloud Fraction, Journal of Climate, 23, 2926–2940, https://doi.org/10.1175/2010JCLI3272.1, http://journals.ametsoc.org/doi/abs/10.1175/2010JCLI3272.1, 2010.

McCoy, D. T., Eastman, R., Hartmann, D. L., and Wood, R.: The change in low cloud cover in a warmed climate inferred from AIRS, MODIS, and ERA-interim, Journal of Climate, 30, 3609–3620, https://doi.org/10.1175/JCLI-D-15-0734.1, 2017.

Medeiros, B., Stevens, B., Held, I. M., Zhao, M., Williamson, D. L., Olson, J. G., and Bretherton, C. S.: Aquaplanets, Climate Sensitivity, and Low Clouds, Journal of Climate, 21, 4974–4991, https://doi.org/10.1175/2008JCLI1995.1, http://journals.ametsoc.org/doi/abs/10.1175/2008JCLI1995.1, 2008.

Muhlbauer, A., McCoy, I. L., and Wood, R.: Climatology of stratocumulus cloud morphologies: microphysical properties and radiative effects, Atmospheric Chemistry and Physics Discussions, 14, 6981–7023, https://doi.org/10.5194/acpd-14-6981-2014, http://www.atmos-chem-phys-discuss.net/14/6981/2014/, 2014.

Myers, T. A. and Norris, J. R.: Observational evidence that enhanced subsidence reduces subtropical marine boundary layer cloudiness, Journal of Climate, 26, 7507–7524, https://doi.org/10.1175/JCLI-D-12-00736.1, 2013.

Natekin, A. and Knoll, A.: Gradient boosting machines, a tutorial, Frontiers in Neurorobotics, 7, https://doi.org/10.3389/fnbot.2013.00021, http://journal.frontiersin.org/article/10.3389/fnbot.2013.00021/abstract, 2013.

Norris, J. R. and Iacobellis, S. F.: North Pacific cloud feedbacks inferred from synoptic-scale dynamic and thermodynamic relationships, Journal of Climate, 18, 4862–4878, https://doi.org/10.1175/JCLI3558.1, 2005.

Painemal, D. and Zuidema, P.: Microphysical variability in southeast Pacific Stratocumulus clouds: synoptic conditions and radiative response, Atmospheric Chemistry and Physics, 10, 6255–6269, https://doi.org/10.5194/acp-10-6255-2010, http://www.atmos-chem-phys.net/10/6255/2010/acp-10-6255-2010.htmlhttp://www.atmos-chem-phys.net/10/6255/2010/, 2010.

Painemal, D., Kato, S., and Minnis, P.: Boundary layer regulation in the southeast Atlantic cloud microphysics during the biomass burning season as seen by the A-train satellite constellation, Journal of Geophysical Research: Atmospheres, 119, 11,288–11,302, https://doi.org/10.1002/2014JD022182, http://doi.wiley.com/10.1002/2014JD022182, 2014.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, https://doi.org/10.1007/s13398-014-0173-7.2, http://dl.acm.org/citation.cfm?id=2078195{%}5Cnhttp://arxiv.org/abs/1201.0490http://arxiv.org/abs/1201.0490, 2011.

Platnick, S. and Twomey, S.: Determining the Susceptibility of Cloud Albedo to Changes in Droplet Concentration with the Advanced Very High Resolution Radiometer, Journal of Applied Meteorology, 33, 334–347, https://doi.org/10.1175/1520-0450(1994)033<0334:DTSOCA>2.0.CO;2, http://journals.ametsoc.org/doi/abs/10.1175/1520-0450{%}281994{%}29033{%}3C0334{%}3ADTSOCA{%}3E2.0.CO{%}3B2, 1994.

Quaas, J., Boucher, O., Bellouin, N., and Kinne, S.: Satellite-based estimate of the direct and indirect aerosol climate forcing, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2007JD008962, http://www.agu.org/pubs/crossref/2008/2007JD008962.shtmlhttp://doi.wiley.com/10.1029/2007JD008962, 2008.

Quaas, J., Stevens, B., Stier, P., and Lohmann, U.: Interpreting the cloud cover – aerosol optical depth relationship found in satellite data using a general circulation model, Atmospheric Chemistry and Physics, 10, 6129–6135, 2010.

Rahn, D. A. and Garreaud, R.: Marine boundary layer over the subtropical southeast Pacific during VOCALS-REx - Part 2: Synoptic variability, Atmospheric Chemistry and Physics, 10, 4507–4519, https://doi.org/10.5194/acp-10-4507-2010, http://www.atmos-chem-phys.net/10/4507/2010/www.atmos-chem-phys.net/10/4507/2010/, 2010.

Sayegh, A., Tate, J. E., and Ropkins, K.: Understanding how roadside concentrations of NOx are influenced by the background levels, traffic density, and meteorological conditions using Boosted Regression Trees, Atmospheric Environment, 127, 163–175, https://doi.org/10.1016/j.atmosenv.2015.12.024, http://dx.doi.org/10.1016/j.atmosenv.2015.12.024, 2016.

Seinfeld, J. H., Bretherton, C., Carslaw, K. S., Coe, H., DeMott, P. J., Dunlea, E. J., Feingold, G., Ghan, S., Guenther, A. B., Kahn, R., Kraucunas, I., Kreidenweis, S. M., Molina, M. J., Nenes, A., Penner, J. E., Prather, K. A., Ramanathan, V., Ramaswamy, V., Rasch, P. J., Ravishankara, A. R., Rosenfeld, D., Stephens, G., and Wood, R.: Improving our fundamental understanding of the role of aerosol-cloud interactions in the climate system, Proceedings of the National Academy of Sciences, 113, 5781–5790, https://doi.org/10.1073/pnas.1514043113, http://www.pnas.org/lookup/doi/10.1073/pnas.1514043113, 2016.

Stier, P.: Limitations of passive remote sensing to constrain global cloud condensation nuclei, Atmospheric Chemistry and Physics, 16, 6595–6607, https://doi.org/10.5194/acp-16-6595-2016, http://www.atmos-chem-phys.net/16/6595/2016/, 2016.

Toniazzo, T., Abel, S. J., Wood, R., Mechoso, C. R., Allen, G., and Shaffrey, L. C.: Large-scale and synoptic meteorology in the south-east Pacific during the observations campaign VOCALS-REx in austral Spring 2008, Atmospheric Chemistry and Physics, 11, 4977–5009, https://doi.org/10.5194/acp-11-4977-2011, http://www.atmos-chem-phys.net/11/4977/2011/, 2011.

Twomey, S.: Pollution and the planetary albedo, Atmospheric Environment (1967), 8, 1251–1256, https://doi.org/10.1016/0004-6981(74)90004-3, http://arxiv.org/abs/1011.1669http://dx.doi.org/10.1088/1751-8113/44/8/085201http://stacks.iop.org/1751-8121/44/i=8/a=085201?key=crossref.abc74c979a75846b3de48a5587bf708fhttp://linkinghub.elsevier.com/retrieve/pii/0004698174900043, 1974.

Várnai, T., Marshak, A., and Yang, W.: Multi-satellite aerosol observations in the vicinity of clouds, Atmospheric Chemistry and Physics, 13, 3899–3908, https://doi.org/10.5194/acp-13-3899-2013, http://www.atmos-chem-phys.net/13/3899/2013/, 2013.

Wilcox, E. M.: Stratocumulus cloud thickening beneath layers of absorbing smoke aerosol, Atmospheric Chemistry and Physics, 10, 11 769–11 777, https://doi.org/10.5194/acp-10-11769-2010, http://www.atmos-chem-phys.net/10/11769/2010/, 2010.

Wood, R.: Stratocumulus Clouds, Monthly Weather Review, 140, 2373–2423, https://doi.org/10.1175/MWR-D-11-00121.1, http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-11-00121.1, 2012.

Wood, R. and Bretherton, C. S.: On the Relationship between Stratiform Low Cloud Cover and Lower-Tropospheric Stability, Journal of Climate, 19, 6425–6432, https://doi.org/10.1175/JCLI3988.1, http://journals.ametsoc.org/doi/abs/10.1175/JCLI3988.1, 2006.

Yamaguchi, T. and Randall, D. A.: Large-Eddy Simulation of Evaporatively Driven Entrainment in Cloud-Topped Mixed Layers, Journal of the Atmospheric Sciences, 65, 1481–1504, https://doi.org/10.1175/2007JAS2438.1, http://journals.ametsoc.org/doi/abs/10.1175/2007JAS2438.1, 2008.

Zhang, Y., Stevens, B., Medeiros, B., and Ghil, M.: Low-Cloud Fraction, Lower-Tropospheric Stability, and Large-Scale Divergence, Journal of Climate, 22, 4827–4844, https://doi.org/10.1175/2009JCLI2891.1, http://journals.ametsoc.org/doi/abs/10.1175/2009JCLI2891.1, 2009.

Zuidema, P., Redemann, J., Haywood, J., Wood, R., Piketh, S., Hipondoka, M., and Formenti, P.: Smoke and Clouds above the Southeast Atlantic: Upcoming Field Campaigns Probe Absorbing Aerosol's Impact on Climate, Bulletin of the American Meteorological Society, 97, 1131–1135, https://doi.org/10.1175/BAMS-D-15-00082.1, http://journals.ametsoc.org/doi/10.1175/BAMS-D-15-00082.1, 2016.