Supplementary Materials: Machine-Learning Based Analysis of Liquid Water Path Adjustments to Aerosol Perturbations in Marine Boundary Layer Clouds Using Satellite Observations

Lukas Zipfel [1,2,*], Hendrik Andersen [1,2] and Jan Cermak [1,2]

1. Supplementary

Spatial gradients are inherent in the LWP and many predictors in the data of the Southeast Atlantic that is used in this work. However, these are not necessarily directly linked to the N_d-LWP relationship. To ensure that the model is not primarily exploiting the spatial gradients to predict the LWP, a secondary data set is created by removing the spatial gradients. The supplementary figures (S2–S5) shown below are created by using this anomalous data set to train the GBRT model.

The spatial gradients are estimated using a linear regression model trained to predict each meteorological variable/cloud state on the basis of the geographical location (latitude-longitude pair) of the observations. The trained model makes a prediction for each meteorological variable/cloud state and for each observation, which is then subtracted from the original data to obtain spatial anomalies (Figure S1). This is done for the predictand (LWP) and all predictors with the exception of PF, as it is based on categorical data.

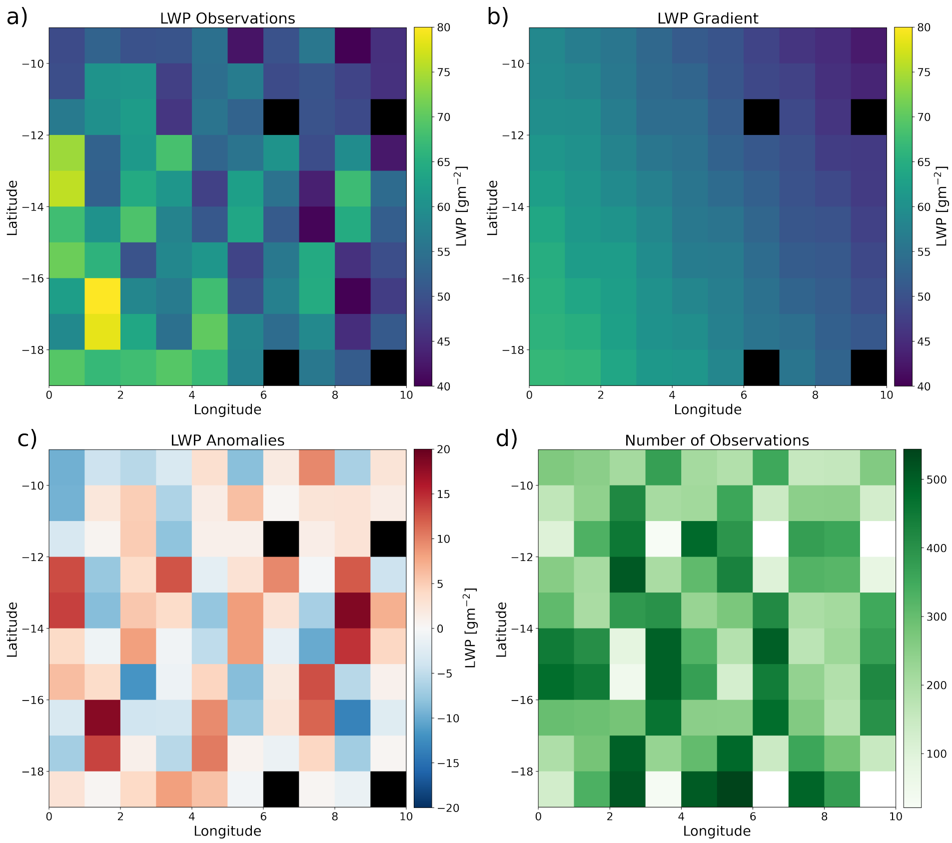

Figure S1. (**a**) The mean AMSR-E LWP in the study area averaged on a 1° x 1° grid, (**b**) the predicted LWP of the linear regression model based on latitude and longitude, and (**c**) the mean LWP anomalies ($(a) - b)$). Pixels with less than 20 data points are displayed in black. Panel (**d**) shows the number of observations for every pixel.

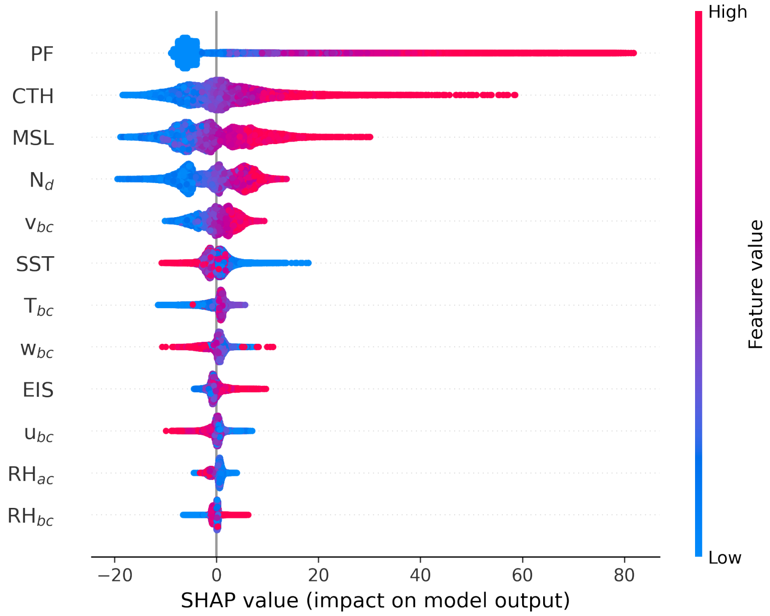

Figure S2. "Beeswarm" plot showing the impact of each feature on the model prediction. Every dot represents an observation with color signifying the original value and the corresponding SHAP value shown on the horizontal axis. Density is made visible by stacking dots on the vertical. The features are sorted by their mean absolute SHAP values in descending order.

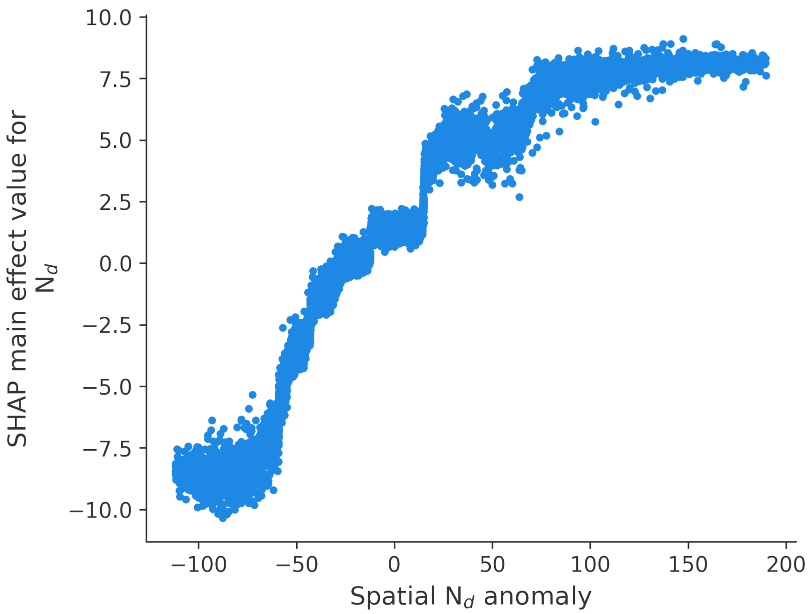

Figure S3. The main effects of N_d anomalies on the prediction of LWP. Main effects show the changes in the model prediction that are solely attributed to the corresponding observed N_d anomaly by removing the interaction effects with other model features from the SHAP values.

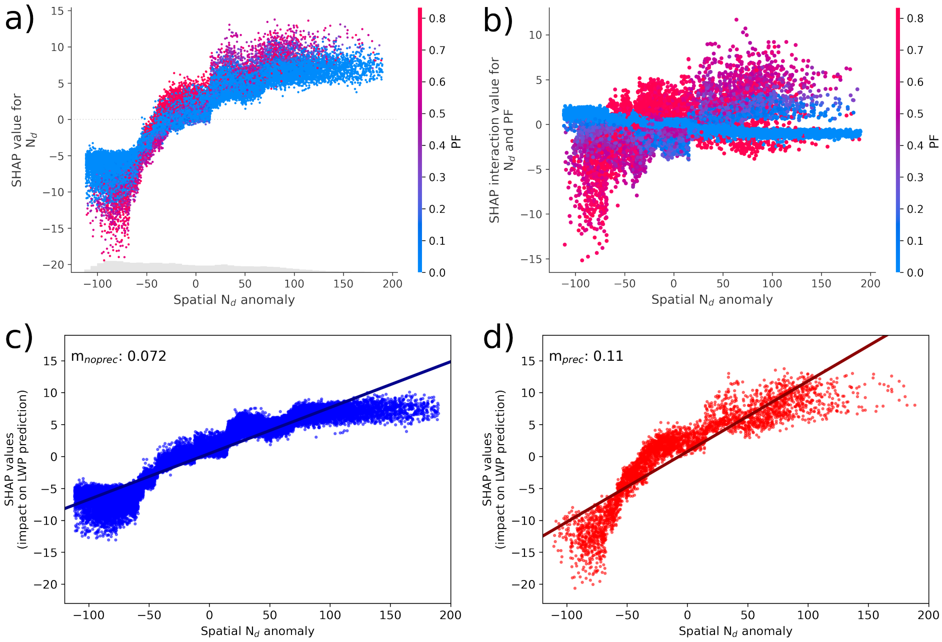

Figure S4. Influence of the PF on the N_d-LWP relationship. SHAP values (**a**) and interaction effects (**b**) for N_d with the color showing the PF. The lower panels show the SHAP values for N_d for non-precipitating situations ($PF = 0$, (**c**)) and for situations where at least 50 % of the cloud groups precipitate ($PF \geq 0.5$, (**d**))

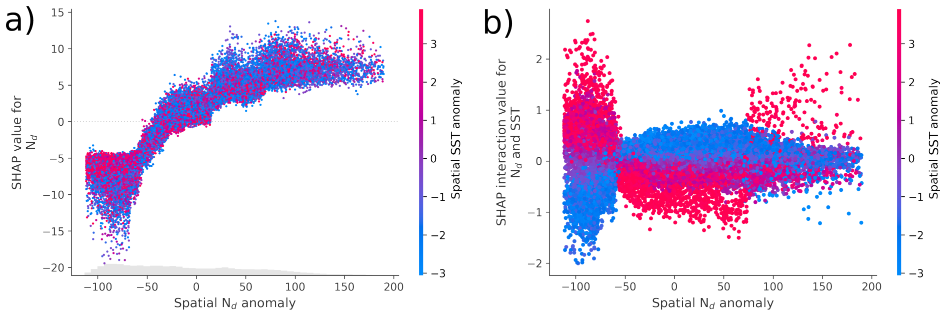

Figure S5. Influence of SST on the N_d-LWP relationship. Panels (**a**) and (**b**) are the same as in Figure S4 with color showing anomalies for SST.