Karlsruhe Institute of Technology

Sharing is encouraged

This presentation participates in OSPP. Outstanding Student & PhD candidate Presentation contest

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821205

FORCeS

ETH zürich

Model vs. observational cloud fraction adjustment using explainable machine learning

Authors: Yichen Jia (contact: Yichen Jia — yichen.jia@kit.edu), David Neubauer, Hendrik Andersen, Ulrike Lohmann, Jan Cermak

1. Introduction

- **Cloud fraction (CLF) adjustment may dominate radiative forcing:** more cloud droplets due to more aerosols lead to increased CLF and thus more cooling.
- One of **the biggest sources of uncertainties in climate models**.
- **Aim: Evaluate** the CLF adjustment for marine boundary layer clouds within the ICOsahedral Non-hydrostatic-Hamburg Aerosol Module (**ICON-HAM**) model.

2. Data and Methods

Daily satellite, reanalysis and ICON-HAM datasets

- N_d (proxy for aerosol) and CLF from Terra satellite and COSP MODIS simulator.
- Meteorological variables from ERA5 reanalysis and ICON-HAM.
- Filtered by cloud temperature > 268 K, effective radius > 4 μm, optical depth > 4. Satellite data: solar and sensor viewing zenith angles < 65° and 55°, respectively.
- Original 1°× 1°grids are aggregated to 5°× 5° "windows". One ML model is trained and tested for a specific window.

Machine learning and SHapley Additive exPlanation (SHAP) values

- **Extreme Gradient Boosting (XGB)** models are used to predict CLF.
- **SHAP values:** contribution of each predictor to each individual model prediction.

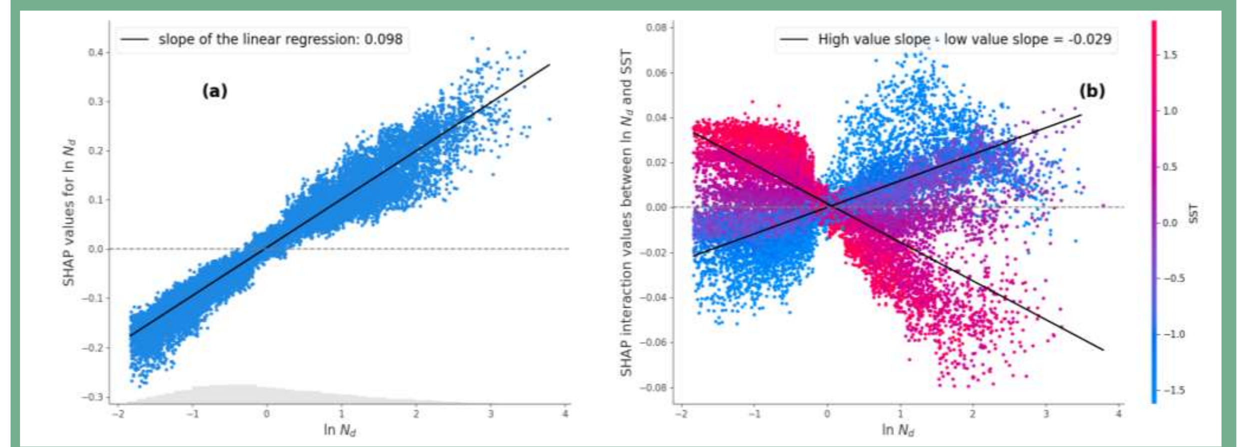

Figure 1: CLF sensitivity to SST (a) and interaction effects between SST and N_d colored by SST (b) in an exemplary 5°× 5° region.

- Figure 1 (a): **CLF sensitivity** is defined as the slope of the linear regression between the SHAP values and feature values of a specific predictor (here N_d).
- Figure 1 (b): **Interaction Index (IAI)** is defined as the slope of linear regressions of the SHAP interaction values and the features values for high SST values (above-average) minus low SST values (below-average).
- **Negative IAI**: sensitivity stronger with low (< mean) meteorological parameters.
- **Positive IAI**: sensitivity stronger with high (> mean) meteorological parameters.

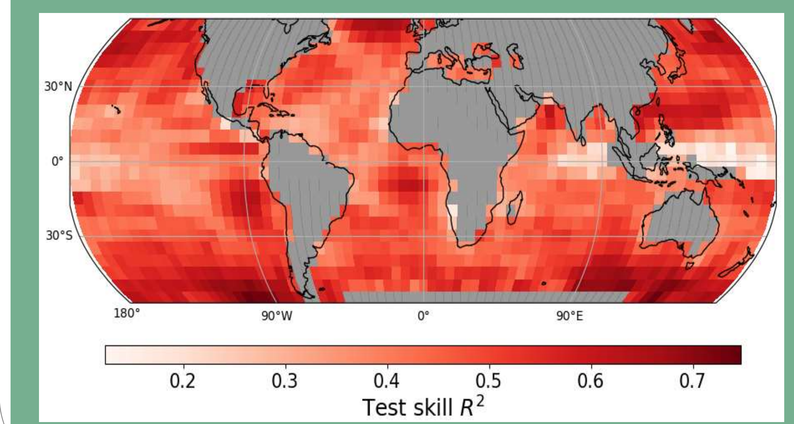

Figure 2: Performance of regional XGB models predicting CLF, global weighted mean ~ 0.45.

3. Results based on satellite and reanalysis (Jia et al., 2023, ACP under review)

CLF sensitivity: global perspective and regional characteristics

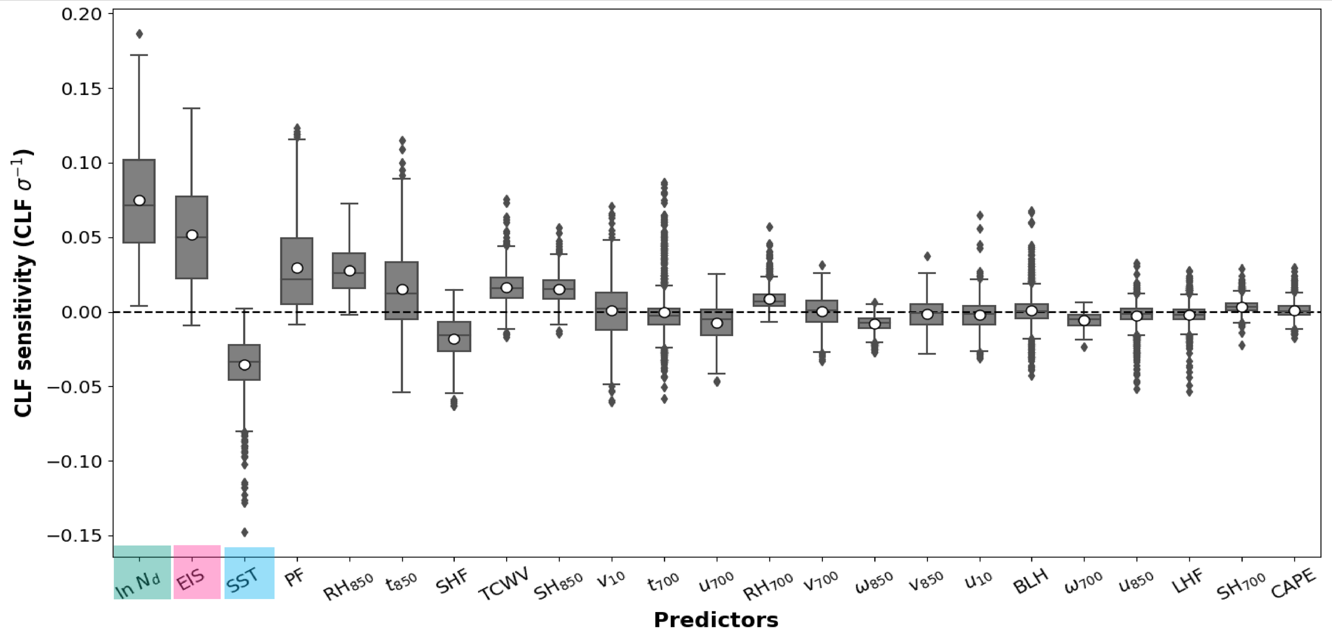

Figure 3: The distribution of the sensitivities of CLF to all predictors. The sequence is sorted descendingly by the mean values of the absolute sensitivitiy values.

- CLF increases with N_d globally.
- The positive sensitivity is prounced in in the regions of frequent stratocumulus to cumulus transition, may be caused by high N_d delaying the transition.
- The positive sensitivity is also marked in the southern hemispheric midlatitudes, should be investigated in future work.

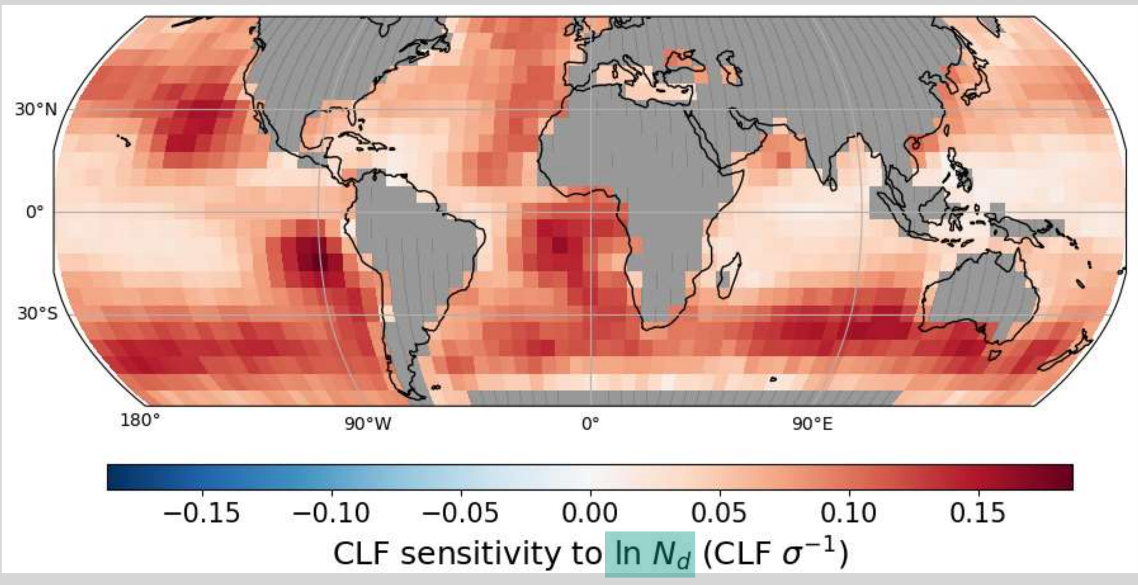

Figure 4: Sensitivity of marine boundary layer cloud fraction to ln N_d.

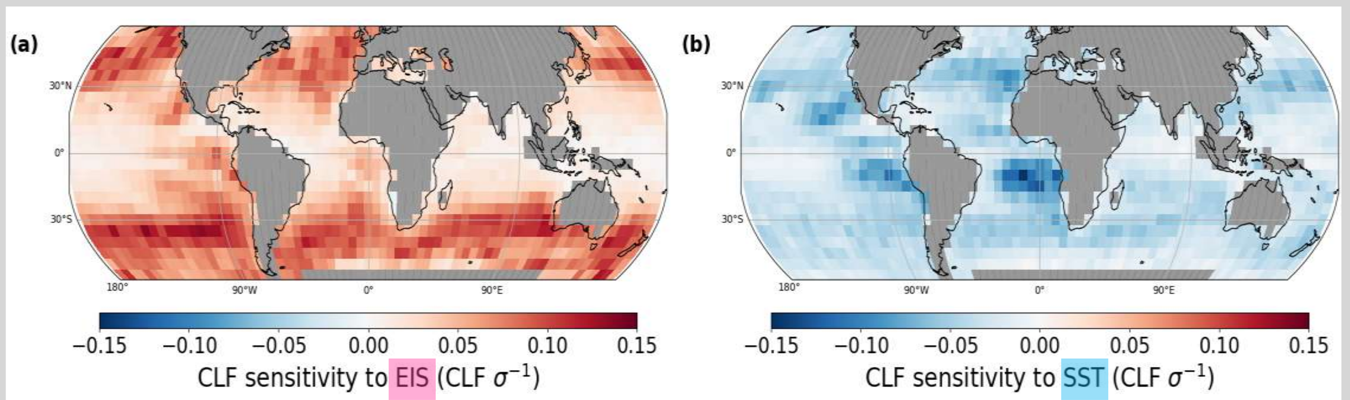

Figure 5: Geographical patterns of CLF sensitivity to estimated inversion strength (EIS) and sea surface temperature (SST).

- CLF is positively associated with EIS, strongest in the stratus and stratocumulus regions.
- CLF is negatively asscoiated with SST globally, strongest in the stratocumulus regions.

Is the N_d–CLF relationship modified by ambient meteorology?

- In general, thermodynamical factors have more interactions with the N_d–CLF sensitivity than dynamical factors.

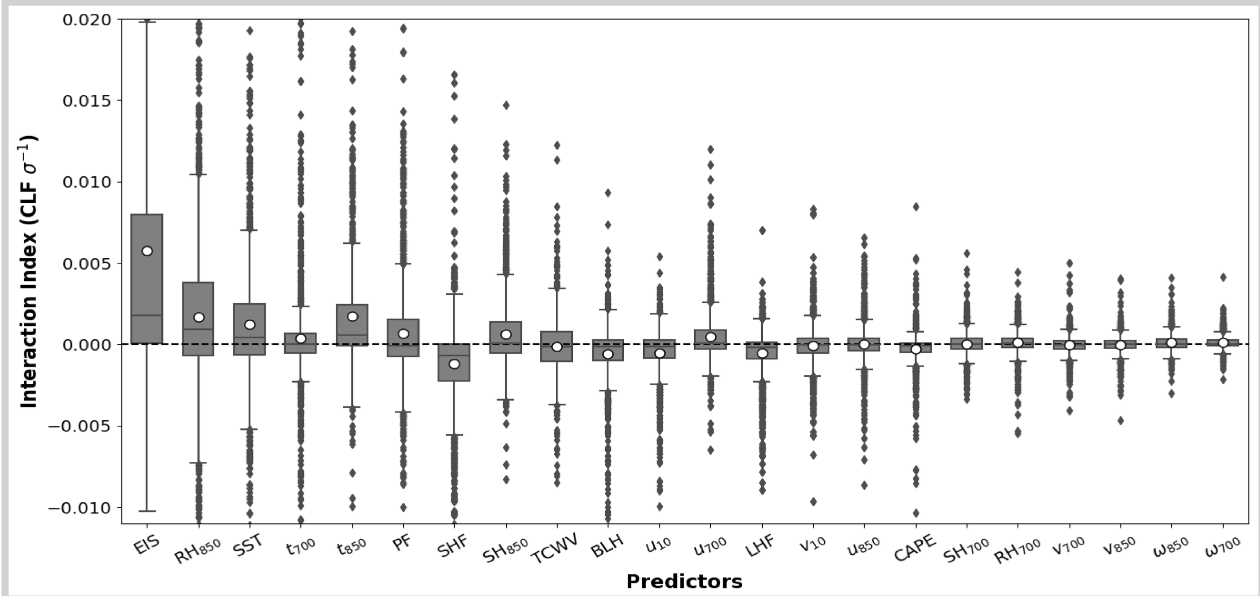

Figure 6: Similar to Fig. 3 but for the Interaction Indices.

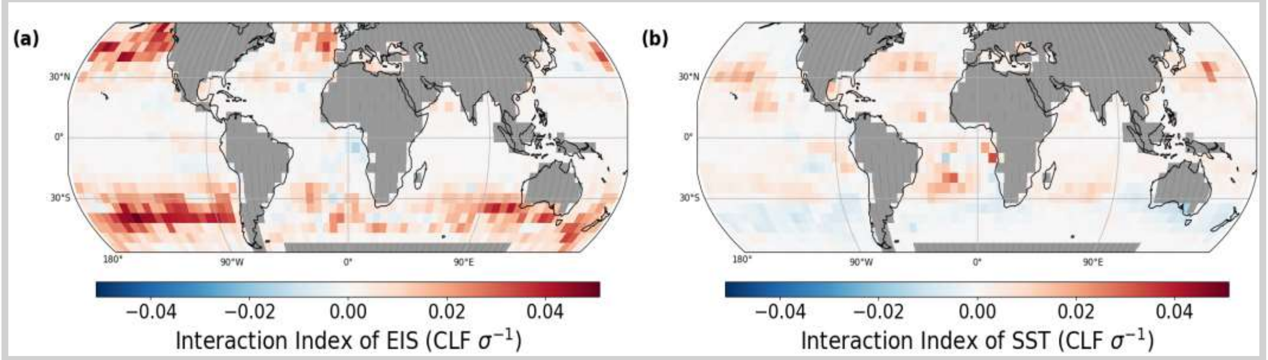

Figure 7: Global patterns of Interaction Indices (IAIs) for the interactive effects between N_d and EIS (a), SST (b).

- EIS exerts positive IAIs over the midlatitudes, reflecting that stronger capping inversions in these regions may amplify the N_d–CLF relationship.
- In the stratocumulus-to-cumulus transition regions, the N_d–CLF sensitivity is stronger with higher SST.

4. Preliminary Results based on ICON-HAM

- Identical XGB and SHAP framework is applied to outputs from ICON-HAM and the COSP MODIS simulator.
- **XGB models perform poorly on the model data.**

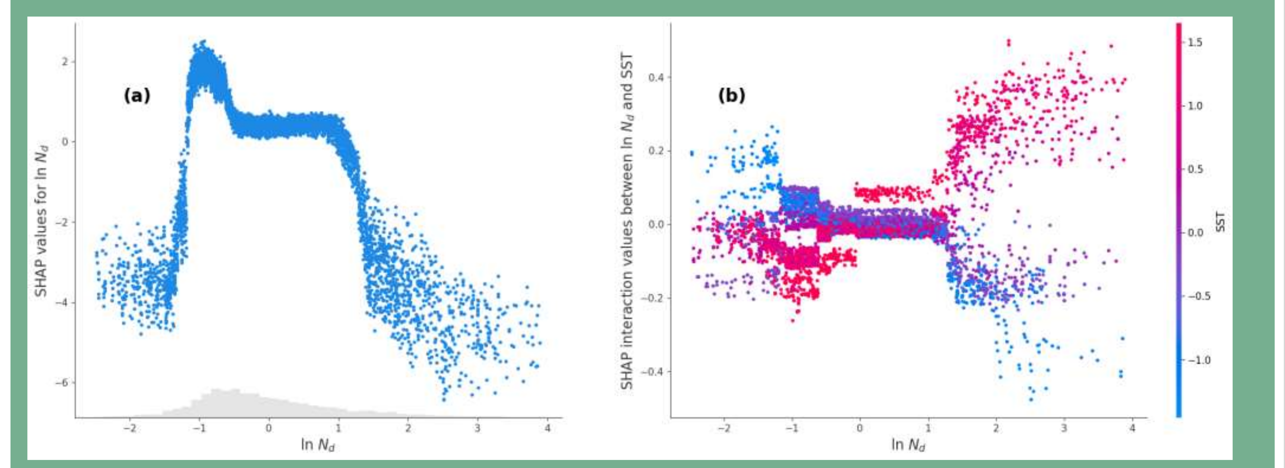

Figure 8: Similar to Fig. 1 for the same 5°× 5° region but using ICON-HAM data.

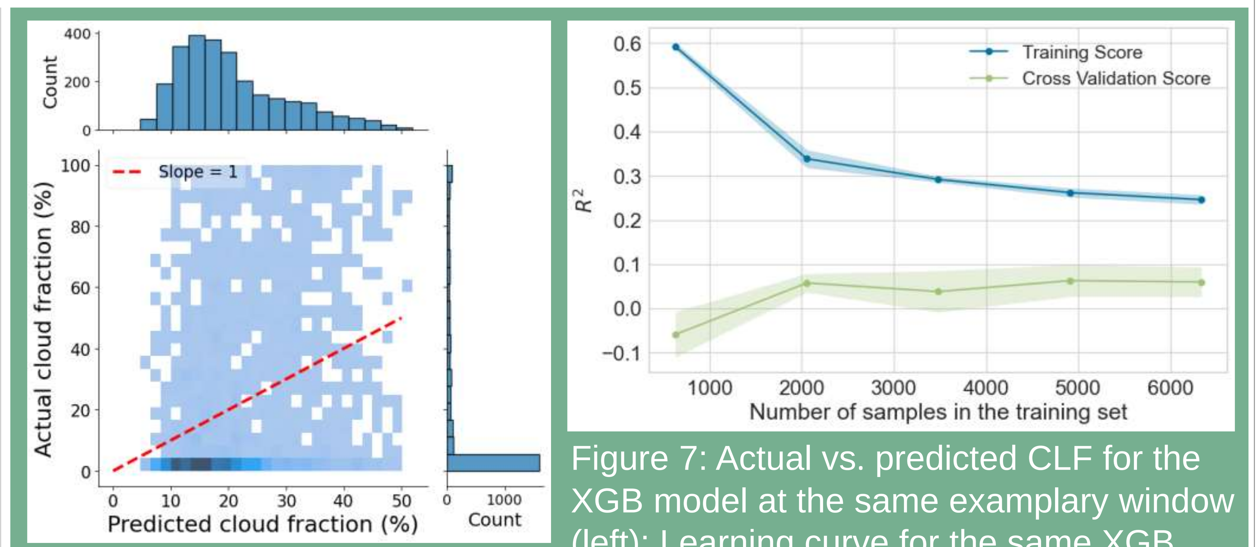

Figure 7: Actual vs. predicted CLF for the XGB model at the same examplary window (left); Learning curve for the same XGB model (right).

- Predicted CLF exhibits very different distribution from the actual CLF values.
- The learning curve shows relatively high bias and high variance.

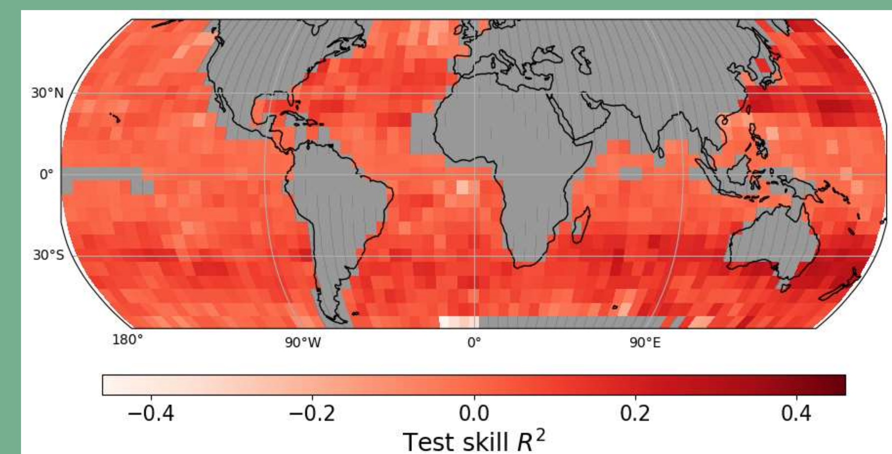

Figure 8: Performance of XGB models. Similar to Fig. 2 but using ICON-HAM data. Global weighted mean ~ 0.047.

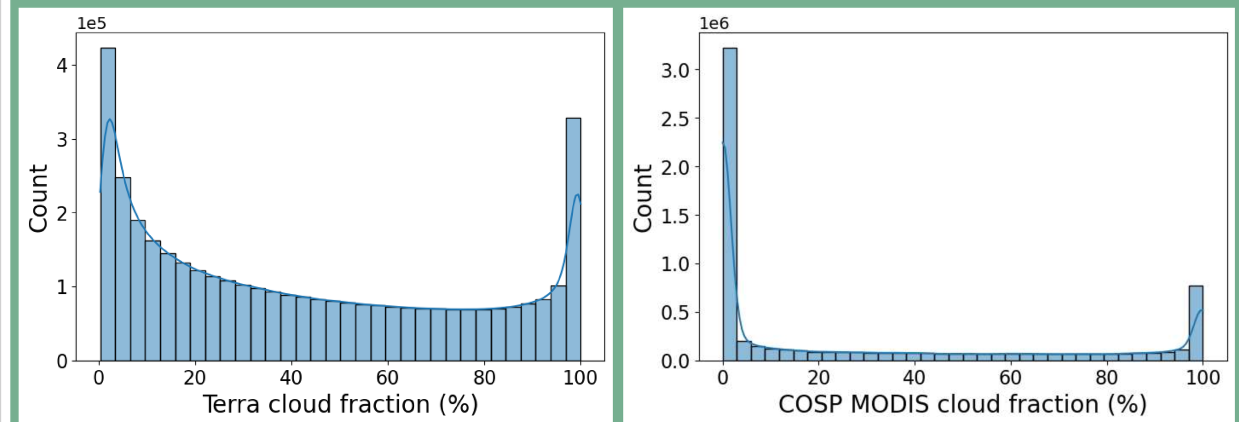

Figure 9: Distributions of cloud fraction values from satellite (left) and COSP MODIS simulator (right).

- Fewer low clouds are simulated with the COSP MODIS simulator than observed by the Terra satellite.

5. Conclusions and Outlook

- The identical XGB + SHAP framework performs poorly with ICON-HAM data.
- The poor performance might be caused by the right-skewed distribution of cloud fraction values from the COSP MODIS simulator.

Future work:

- Identify the cause for the above discrepancy.
- Improve the machine learning models accordingly and apply the same framework to both observational and model datasets.
- Compare sensitivities and interactive effets with meteorological conditions quantified by SHAP approach.

Study based on satellite and reanalysis data:
Jia, Y., Andersen, H., and Cermak, J.: Analysis of the cloud fraction adjustment to aerosols and its dependence on meteorological controls using explainable machine learning, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-1667, 2023.

Scan for abstract