Geophysical Research Letters

RESEARCH LETTER
10.1029/2024GL109735

Improving Low-Cloud Fraction Prediction Through Machine Learning

Haipeng Zhang[1,2], **Youtong Zheng**[3,4], **and Zhanqing Li**[1,2]

[1]Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD, USA, [2]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA, [3]Department of Earth and Atmospheric Science, University of Houston, Houston, TX, USA, [4]Institute of Climate and Atmospheric Science, University of Houston, Houston, TX, USA

Key Points:

- Machine learning (ML) models substantially improve "too few" low-cloud problems in the subtropics compared to traditional cloud schemes
- They also show marked improvement in representing low-cloud fraction (LCF) variations during the stratocumulus-to-cumulus transition
- Including the effect of moisture source in ML models is crucial to representing spatiotemporal variations in LCF in the midlatitudes

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Z. Li and Y. Zheng,
zhanqing@umd.edu;
yzheng18@uh.edu

Citation:

Zhang, H., Zheng, Y., & Li, Z. (2024). Improving low-cloud fraction prediction through machine learning. *Geophysical Research Letters*, *51*, e2024GL109735. https://doi.org/10.1029/2024GL109735

Received 11 NOV 2023
Accepted 20 JUL 2024

Abstract In this study, we evaluated the performance of machine learning (ML) models (XGBoost) in predicting low-cloud fraction (LCF), compared to two generations of the community atmospheric model (CAM5 and CAM6) and ERA5 reanalysis data, each having a different cloud scheme. ML models show a substantial enhancement in predicting LCF regarding root mean squared errors and correlation coefficients. The good performance is consistent across the full spectrums of atmospheric stability and large-scale vertical velocity. Employing an explainable ML approach, we revealed the importance of including the amount of available moisture in ML models for representing spatiotemporal variations in LCF in the midlatitudes. Also, ML models demonstrated marked improvement in capturing the LCF variations during the stratocumulus-to-cumulus transition (SCT). This study suggests ML models' great potential to address the longstanding issues of "too few" low clouds and "too rapid" SCT in global climate models.

Plain Language Summary Low clouds impose a strong radiative cooling effect on Earth's climate. Predicting low-cloud fraction (LCF) is, however, challenging in global climate models (GCMs), partly due to some deficiencies in cloud parameterization schemes. Machine learning (ML) models might fill this gap as it is recognized as an efficient, economical, and accurate method to make predictions. In this study, we find that ML models (XGBoost) exhibit superior proficiency in predicting LCF regarding root mean squared errors and correlation coefficients compared to two generations of the community atmospheric model (CAM5 and CAM6) and ERA5 reanalysis data, each having a different cloud scheme. This improvement helps address one important issue of "too few" low clouds in GCMs. Furthermore, ML models demonstrate marked improvement in representing LCF variations when stratocumulus clouds transition to cumulus clouds, as opposed to too rapid decreases in LCF simulated by two CAMs and ERA5. Such findings testify to the unique role of ML models in refining the parameterization of LCF within GCMs.

1. Introduction

Low clouds have a substantial cooling effect on Earth's climate. However, due to their poor representation in global climate models (GCMs), the response of cloud properties (e.g., low-cloud fraction or LCF) to global warming is a major source of uncertainty in predicting future climate (Bony & Dufresne, 2005; Soden & Vecchi, 2011; Vial et al., 2013; H. Zhang et al., 2018).

The parameterization of clouds in GCMs is a complex and critical aspect of climate modeling (Edwards, 2011), as clouds are controlled by various physical processes, encompassing small-scale turbulence, mesoscale circulations, radiation, and microphysics (Bretherton, 2015; Ceppi et al., 2017; Houze, 2014; Wood, 2012; M. Zhang et al., 2013). Simulating cloud fraction accurately is one of the most challenging aspects of cloud parameterization. Currently, GCMs employ two main types of schemes for parameterizing cloud fraction: the prognostic scheme and the diagnostic scheme. The prognostic cloud scheme calculates the temporal variation of cloud fraction based on source and sink terms associated with advection, cumulus convection, stratiform condensation, boundary layer turbulence, and dissipation caused by evaporation and precipitation. Typical examples of prognostic cloud schemes are the Tiedtke scheme (Tiedtke, 1993) employed in the ECMWF Integrated Forecast System (IFS) (ECMWF, 2016) and the Geophysical Fluid Dynamics Laboratory's atmospheric model (Donner et al., 2011; Zhao et al., 2018). In contrast, the diagnostic scheme calculates the cloud fraction using the instant grid-mean atmospheric properties like relative humidity and/or cloud water contents, such as the Xu-Randall scheme (Xu & Randall, 1996). Another more advanced cloud scheme, CLUBB that unifies the representations

of cloud macrophysics, boundary-layer turbulence, and shallow cumulus (Golaz et al., 2002a, 2002b; Larson & Golaz, 2005; Larson et al., 2002, 2012), diagnoses cloud fraction by analytic integration over the probability density function (PDF) of cloud liquid water mixing ratio (Larson, 2017).

However, both types of cloud schemes demonstrate similar deficiencies in simulating low clouds in GCMs. GCMs tend to underestimate LCF and overestimate cloud optical depth, a problem commonly known as the "too few, too bright" low-cloud problem (e.g., Nam et al., 2012). The LCF underestimation is usually found over the Southeast Pacific (Jian et al., 2021), the Southern Ocean (Schuddeboom & McDonald, 2021), and East Asia (G. Chen et al., 2022; Z. Chen et al., 2022). Another challenging issue of cloud simulations is the too-rapid stratocumulus-to-cumulus transition (SCT) in the subtropics and midlatitude regions (Bogenschutz et al., 2013; H. Zhang et al., 2024a), manifested by a rapid reduction in LCF along the transition trajectory.

Recently, machine learning (ML) algorithms have been demonstrated to be an efficient, accurate, and economical approach to predicting or diagnosing geophysical variables, such as ice sheet melt (Sellevold & Vizcaino, 2021), seasonal precipitation (Pan et al., 2022), and pollution concentrations (J. Li et al., 2022). Reichstein et al. (2019) argued that ML can be a promising tool to help predict cloud amount and explore cloud physics processes. Therefore, ML models might establish a novel LCF parameterization scheme through various resolved meteorological factors (MFs) to bridge the aforementioned gaps. However, a comprehensive examination of ML in predicting LCF, particularly on shorter timescales (e.g., sub-daily), has yet to be done. We shall examine the performance of ML models in representing LCF against traditional physics-based cloud schemes (e.g., Tiedtke and CLUBB). Specifically, we focus on investigating whether ML models can improve the "too few" low-cloud problem and the overly rapid SCT issue. Moreover, the relative contributions of various MFs on LCF across different regions will be sorted out through an explainable ML approach called SHapley Additive exPlanations (SHAP; Lundberg & Lee, 2017; Lundberg et al., 2018), helping enhance our understanding of meteorological controls on clouds.

2. Methodology

2.1. Satellite and Reanalysis Data

We obtain LCF from the CERES SYN Ed4 product that provides satellite-based estimation of low-level cloud amount (1° latitude by 1° longitude resolved hourly) throughout 2004–2013. This product utilizes 16 geostationary satellites. To ensure consistency across the various satellite sensors, the instrument channels are cross-calibrated to match the cloud retrievals from the MODIS collection 5 product (Doelling et al., 2013). The fifth-generation ECMWF reanalysis data (ERA5; Hersbach et al., 2020) provide hourly meteorological variables such as horizontal wind fields, temperature, and moisture in the lower free troposphere, with a grid resolution of $0.25° \times 0.25°$. We linearly interpolate the ERA5 data to 1-degree grids to match the cloud product's resolution. Utilizing a data set spanning 10 years with an hourly temporal resolution entails significant computational expense. Consequently, a six-hourly temporal resolution is adopted as an alternative, providing sufficient accuracy for trajectory calculations within the Lagrangian framework evaluation. In this study, we primarily focus on marine warm clouds. The regions of interest are the oceans between 60°S and 60°N.

2.2. ML Models and Predictor Selection

The ML algorithm employed in this study is the eXtreme Gradient Boosting Decision Tree (XGBoost; T. Chen & Guestrin, 2016), which is a scalable and accurate implementation of gradient boosting, offering improved speed over traditional boosted tree models. We have chosen the XGBoost algorithm instead of opting for more complex models like densely connected feedforward neural networks, mainly due to its relative simplicity, adequate accuracy, and interpretability. Two XGBoost models are built based on two different sets of predictors. We generally follow Y. Chen et al. (2021) to select the first set of predictors. The basic idea is to use the coarse-grid information of thermodynamic and dynamic properties of the low-level troposphere along with atmospheric column features and surface forcing to predict LCF, of which 10 MFs are chosen as they show a persistently significant role in Y. Chen et al. (2021)'s random-predictor-combined experiments. The 10 MFs comprise the following variables: potential temperature at 1,000, 850, and 700 hPa (θ_{1000}, θ_{850}, θ_{700}); relative humidity at 1,000, 850, and 700 hPa (RH_{1000}, RH_{850}, RH_{700}); horizontal wind speed at 1,000 hPa (U_{1000}); vertical velocity at

700 hPa (ω_{700}); latent heat flux (LHF); and column-integrated precipitable water vapor (PWV), as summarized in Table S1 in Supporting Information S1.

As a comparison, the second set of predictors originates from the seven-variable heuristic model developed by Qu et al. (2015). Here, we replace the estimated inversion strength (EIS; Wood & Bretherton, 2006) with the lower-tropospheric stability (LTS; Klein & Hartmann, 1993) in the original Qu's model, as we intend to keep each predictor as simple as possible (i.e., using a more basic MF that is defined without any assumption) to avoid the prediction errors induced by a predictor itself. These seven MFs consist of LTS, LHF, Δq, ω_{700}, RH_{700}, U_{1000}, and T_{adv}, where Δq is the specific humidity difference between 1,000 and 700 hPa and T_{adv} is horizontal temperature advection (summarized in Table S1 in Supporting Information S1). Each predictor in Qu's model explicitly represents a physical process. For example, LTS or EIS represents the influence of cloud top inversion strength (Bretherton et al., 2013), and latent heat fluxes (LHF) show the impact of surface-flux-induced entrainment drying (Bretherton & Blossey, 2014), with others detailed in Qu et al. (2015). Per the number of predictors, these two XGBoost models are named XGB10 and XGB7, respectively. XGB7 features the explicit representation of seven physical processes that control clouds, while XGB10 can implicitly represent more physical processes by a random combination of those basic thermodynamic and dynamic properties. This setup of gradation in predictor complexity helps understand the role of the well-known physical processes and those beyond in the accuracy of LCF representations.

For these two ML models, the gridded data set from 2006 to 2013 is used for training, and the rest (2004 and 2005) is for testing (independent training/test split about 80%/20%). Total sample sizes for training and testing are around 322,000,000, and 80,800,000 respectively. All input features are standardized by removing their means and scaling them to unit variance, and LCF is linearly transformed between 0 and 1. Details on building XGB10 and XGB7 can be found in Text S1 in Supporting Information S1. Prediction results from the test set are used for analysis in this study.

2.3. GCMs and Nudging Experiments

LCF predictions from ML models will be compared against outputs from two GCMs and ERA5 reanalysis data, each having distinct cloud schemes. The two models adopted are the Community Atmosphere Model version 5 (CAM5; Hurrell et al., 2013) and version 6 (CAM6; Danabasoglu et al., 2020). We refer to the models CAM5 and CAM6 collectively as "the CAMs" throughout the paper. In early versions of the CAMs, marine stratocumulus clouds and shallow cumulus clouds are separately parameterized: stratocumulus clouds are diagnosed using empirical relationships between stratocumulus cloud fraction and LTS, while shallow cumulus cloud fraction is linked to updraft mass flux (Neale et al., 2010). In CAM5, stratus clouds are more finely represented by diagnosing separate liquid and ice stratus fractions as opposed to single-phase stratus fraction in CAM3 and CAM4 (Neale et al., 2012). The recent CAM (CAM6) has implemented a more advanced cloud scheme, called the Cloud Layers Unified by Binormals (CLUBB; Golaz et al., 2002a, 2002b; Larson & Golaz, 2005; Larson et al., 2002, 2012), which integrated the shallow convection and boundary layer clouds into one unified framework. This unified scheme has been shown to significantly improve the simulations of low clouds, especially over the SCT regime in CAM5 (Bogenschutz et al., 2013).

To compare with ML models, the default CAM5 and CAM6 were used to perform AMIP simulations (i.e., the FHIST component set) with specified sea surface temperatures and sea ice. They were run with a horizontal resolution of 0.9° × 1.25°, up to 32 vertical layers. The models were initialized on 1 January 2004, with ERA5 interpolated onto its grid and integrated for 2 years. To minimize the influence of simulation errors from large-scale meteorological fields on LCF predictions, two CAMs were continuously nudged to ERA5 winds, temperature, and moisture at every time step (30 min) during the 2-year integrations. More specifically, the target state was updated to ERA5 at the top of every 6 hr and to ERA5 linearly interpolated at other time steps. The nudging simulations were completed with a constant nudging time scale of 1 hr. Instantaneous data were output every 6 hr for analysis. Figure S1 in Supporting Information S1 compares surface winds between the nudging outputs and ERA5, highlighting a close resemblance with small bias errors and root mean square errors (mostly <2 m/s) and large correlation coefficients (approximately 1). Also, comparisons of clouds between model output and satellite observations are hampered by their inherent differences. For instance, low clouds are obscured by high or middle clouds in MODIS, but without such a restriction in the model world. For a lucid comparison, a

satellite simulator, the Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package (COSP; Bodas-Salcedo et al., 2011) for MODIS was enabled for the CAM cloud fraction.

Besides the CAMs, LCF from ERA5 is also used for comparison. ERA5 is produced using 4D-Var data assimilation and model forecasts in CY41R2 of the ECMWF IFS (ECMWF, 2016), in which cloud fraction is prognosed by the Tiedtke scheme (Tiedtke, 1993). ERA5 LCF can be, therefore, regarded as outputs from a GCM with realistic large-scale meteorology specified. Different from CAM outputs, the COSP-simulated cloud fraction is unavailable for ERA5. Even without employing COSP, the comparison of ERA5 LCF with satellite observations remains decent in certain regions that lack high clouds. This is demonstrated by Figure S2 in Supporting Information S1, which compares the COSP-simulated LCF with the non-COSP output LCF in the CAMs, showing a notable similarity in subsidence regions, particularly where stratocumulus clouds dominate. The modeled LCF is determined as the maximum cloud fraction below 700 hPa.

2.4. SHAP Explainability Analysis

To quantify the individual contribution of each predictor to LCF predictions, the SHAP (SHapley Additive exPlanations; Lundberg & Lee, 2017; Lundberg et al., 2018) explainable ML analysis is employed to interpret the XGBoost model behavior. This statistical approach is based on coalitional game theory, which calculates the contribution of a predictor as the difference between the predictions of an ML model in the presence and absence of this specific predictor for all possible predictor combinations. The SHAP value denotes the contribution of each predictor to the prediction. More details on SHAP analysis can be found in Text S2 in Supporting Information S1.

2.5. Lagrangian Frameworks

To examine the performance of models in simulating the SCT, a novel Lagrangian framework (Christensen et al., 2023; Eastman et al., 2021; Eastman & Wood, 2018) is adopted, which samples the same parcels as they evolve in time and space. This framework allows evaluating variations in LCF along the trajectories on timescales of hours to days. Within the framework, numerous Lagrangian trajectories are calculated from ERA5 (used to track ERA5 LCF, ML-predicted LCF, and observed LCF) and CAM5/CAM6 outputs (used to track their own simulated LCF), which are driven by their isobaric winds at 925 hPa, assuming that this level approximates the cloud level within the boundary layer (Eastman & Wood, 2018). We focus on four regions where the SCT prevalently occurs (following Eastman and Wood (2018)): Northeast Pacific (−155° to −115°E, 15° to 30°N), Southeast Pacific (−105° to −70°E, −30° to −5°N), Southeast Atlantic (−15° to 15°E, −30° to −5°N), and East Indian (62.5° to 112.5°E, −30° to −20°N). In these regions, starting points of trajectories are sampled at an equal distance of 200 km (see Figure S3 in Supporting Information S1). For each starting point, the 36-hr forward trajectories are computed every 6 hr for the years 2004–2005 (an example is illustrated in Figure S4 in Supporting Information S1). In this manner, we achieve an even temporal and spatial sampling of all types of trajectories under various dynamic conditions. To select SCT cases, we choose those trajectories experiencing cold-advection (CADV) conditions and starting with LCF greater than 0.9, according to the SCT's characteristics (Albrecht et al., 1995; Bretherton et al., 1999; Wyant et al., 1997; H. Zhang et al., 2023; Zheng, Zhang, Rosenfeld, et al., 2021). CADV is defined as a situation where the linearized temporal rate of SSTs over 36 hr along a trajectory is negative at the significance level of 0.05. LCF is tracked along the selected trajectories for evaluation. It is notable that LCF variations are examined through the ensemble behavior of thousands of trajectories (ensemble average) rather than focusing on a single trajectory.

3. Results

3.1. Evaluation of Global LCF Predictions

Figure 1 compares the global distribution of climatological mean LCF for the years 2004–2005 between different models and satellite observations. Both CAMs (with COSP enabled) underestimate LCF in tropical oceans, particularly the stratocumulus regions near the west coasts of continents, marked by the black boxes (Figures 1a–1c), which has been a common problem in many GCMs (Vignesh et al., 2020). This “too few” low-cloud problem becomes more pronounced in the new generation model, CAM6, in which the low-bias areas are expanded much more from the continents to deep oceans compared to CAM5. Furthermore, CAM6 fails to capture the peak values of LCF observed near the west coasts of California, Peru, and Australia in the subtropics. This deficiency in CAM6 is likely due to the assumed equal width double-Gaussian PDF of vertical velocity in CLUBB, which

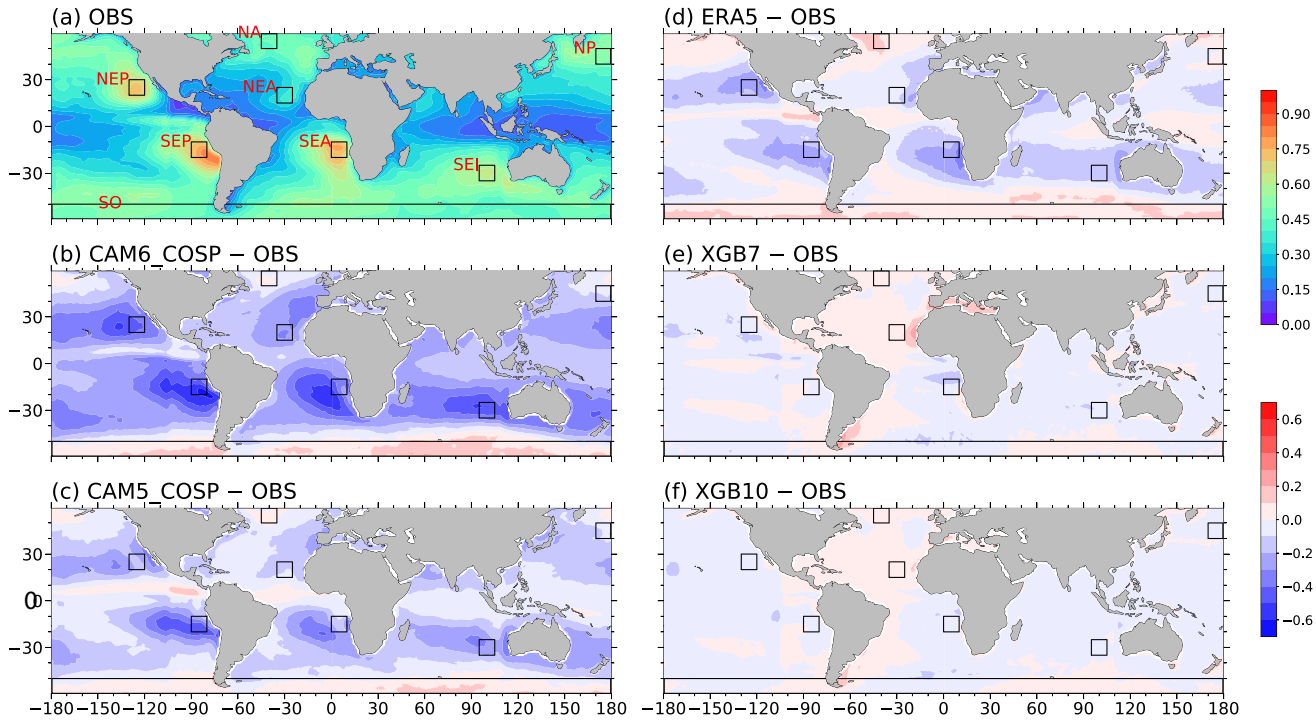

Figure 1. (a) Global maps of climatological mean low-cloud fraction (LCF) for the years 2004–2005 from CERES SYN Ed4 product. Other panels show the climatological mean difference in LCF between a model and the CERES product: (b) CAM6 with COSP enabled, (c) CAM5 with COSP enabled, (d) ERA5, (e) XGB7, and (f) XGB10. The black boxes of 10° × 10° mark the eight selected regions where stratocumulus decks are prevalent: North Atlantic (NA), North Pacific (NP), Northeast Pacific (NEP), Northeast Atlantic (NEA), Southeast Pacific (SEP), Southeast Atlantic (SEA), Southeast Indian Ocean (SEI), and Southern Ocean (SO), identified by Klein and Hartmann (1993).

yields relatively higher skewness of vertical velocity (indicating asymmetric vertical motions) in stratocumulus regions, causing significantly low biases in LCFs (T. Li et al., 2022). Compared to the CAMs, the issue of "too few" low clouds has been greatly improved in ERA5 across the majority of tropical regions. However, the underestimation of LCF persists in stratocumulus regions (Figure 1d). In contrast, two ML models, XGB7 and XGB10, exhibit significant enhancements in addressing these deficiencies within the CAMs and ERA5 (the enhancement in each ML model passed the two-tailed *t*-tests at the 0.05 significant level; Figures 1e–1f vs. Figures 1b–1d). They effectively capture the peak values of LCF in stratocumulus regions, along with the spatial patterns of LCF bearing a close resemblance to those observed.

To further explore the prediction skills of LCF, three error metrics (bias error, root mean square error (RMSE), and correlation coefficient (r)) are presented in Figures 2a–2c, respectively. The overall performance of LCF predictions in terms of these three metrics is on par among the CAMs and ERA5. The "too few" low-cloud problems in CAM5 and CAM6 are confirmed by notably negative biases in Figure 2a, especially in stratocumulus regions (Figure 2d). Although ERA5 more accurately captures observed climatological LCF patterns in the tropics compared to the CAMs, its overall skill globally in terms of r becomes poorer (Figure 2c). This decline is primarily attributed to the biases in the midlatitudes because no COSP is employed (see discussion in Section 2.3). ML models exhibit superior performance relative to the CAMs and ERA5 regarding all the metrics. Moreover, in comparison to XGB7, XGB10 demonstrates a significant increase in r, indicating that the control of spatio-temporal variations in LCF involves complex physical processes not fully encapsulated by the seven well-known processes summarized in Qu et al. (2015).

To discern the efficacy and reliability of these models across cloud regimes and dynamic regimes, we examine the dependencies of error metrics on LTS (Figures 2d–2f) and large-scale vertical velocity (Figures 2g–2i). A threshold value of LTS greater than 18.5 K is employed to distinguish stratocumulus regimes, whereas a value smaller than 15.4 K is used to identify cumulus regimes, with the range between these thresholds characterizing the transition clouds, adapted from Medeiros and Stevens (2011). ML models exhibit consistently good skills in

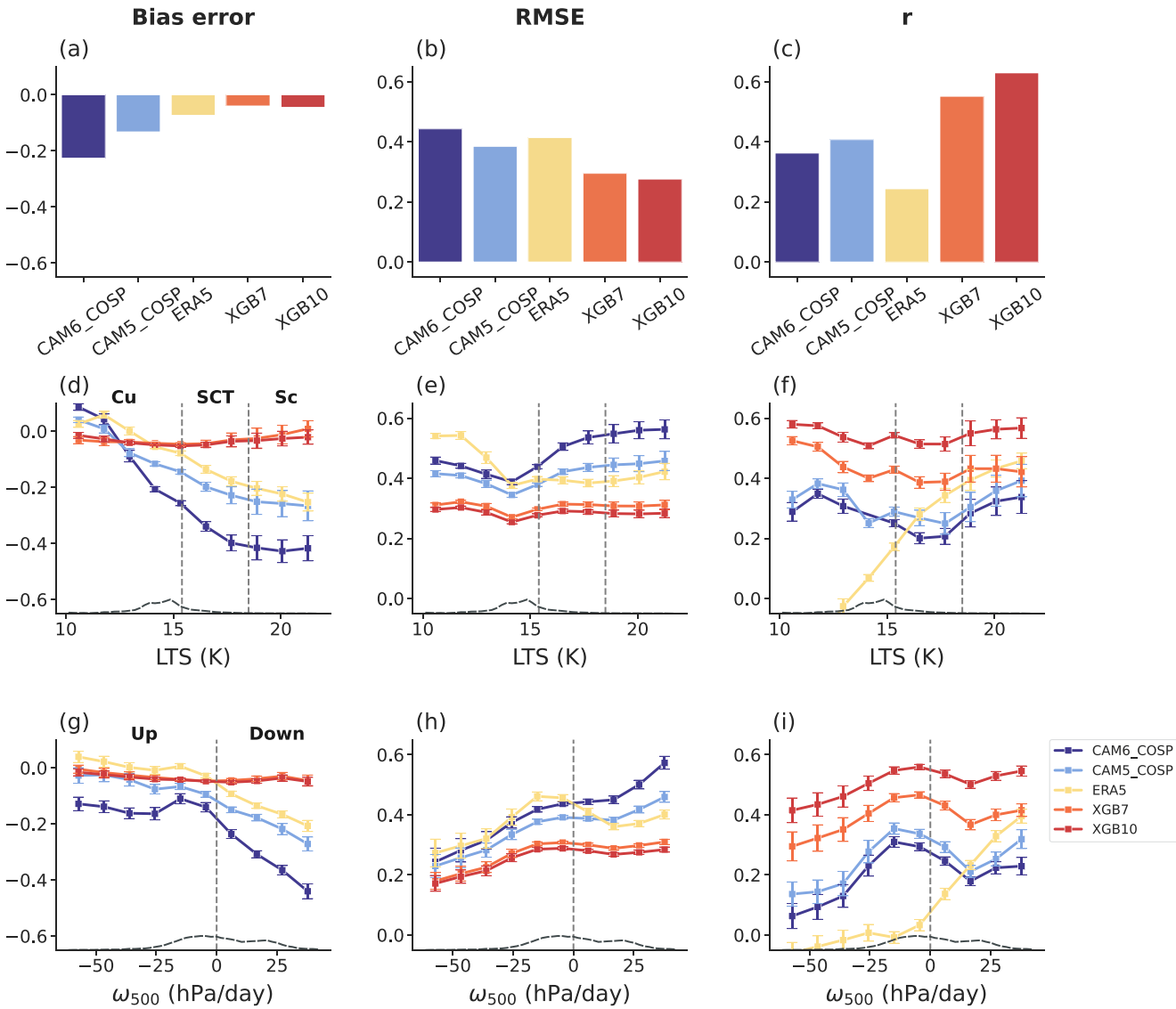

Figure 2. The comparison of the performance of low-cloud fraction (LCF) prediction among different models. (a) The bias error (model minus observation), (b) the root mean squared error (RMSE), and (c) the correlation coefficients (r) of predicted LCF compared against observed LCF for all the cases. The r values shown in (c) are all statistically significant at the 0.05 level. Panels (d)–(f) show the dependence of the bias error, RMSE, and r on lower tropospheric stability (LTS), respectively. Panels (g)–(i) show the same but for the dependence on large-scale vertical velocity at 500 hPa (ω_{500}). The error bars represent the expanded standard error of 10 for visualization. The dashed gray lines show the probability density function (PDF) of the corresponding meteorological factor. Three cloud regimes are differentiated according to the empirical thresholds of LTS: cumulus (Cu), stratocumulus-to-cumulus (SCT), and stratocumulus (Sc) regimes, as marked in panels (d)–(f) so are updraft marked in panels (g)–(i).

predicting LCF across cloud regimes, as opposed to notably poorer and more variable performance by traditional cloud schemes in the CAMs and ERA5. Specifically, in the CAMs, the bias error and RMSE generally deteriorate with increasing atmospheric stability, while a "V" shape is observed for r with valley observed in the transition regimes. The robust skills of ML models are further reinforced by their minimal susceptibility to dynamic regime variations compared to the CAMs and ERA5. Due to the limitation of no COSP in ERA5, its prediction skills of spatiotemporal LCF variations drop to zero in cumulus regimes or updraft regimes where high-level clouds potentially exist. But in stratocumulus regimes, ERA5 outperforms the CAMs.

3.2. Evaluation of Regional LCF Predictions

To gain a deeper insight into the modeled LCF, we delve into the regional differences of LCF simulation errors in eight stratocumulus-dominated regions (see black boxes in Figure 1a): North Atlantic (NA), North Pacific (NP),

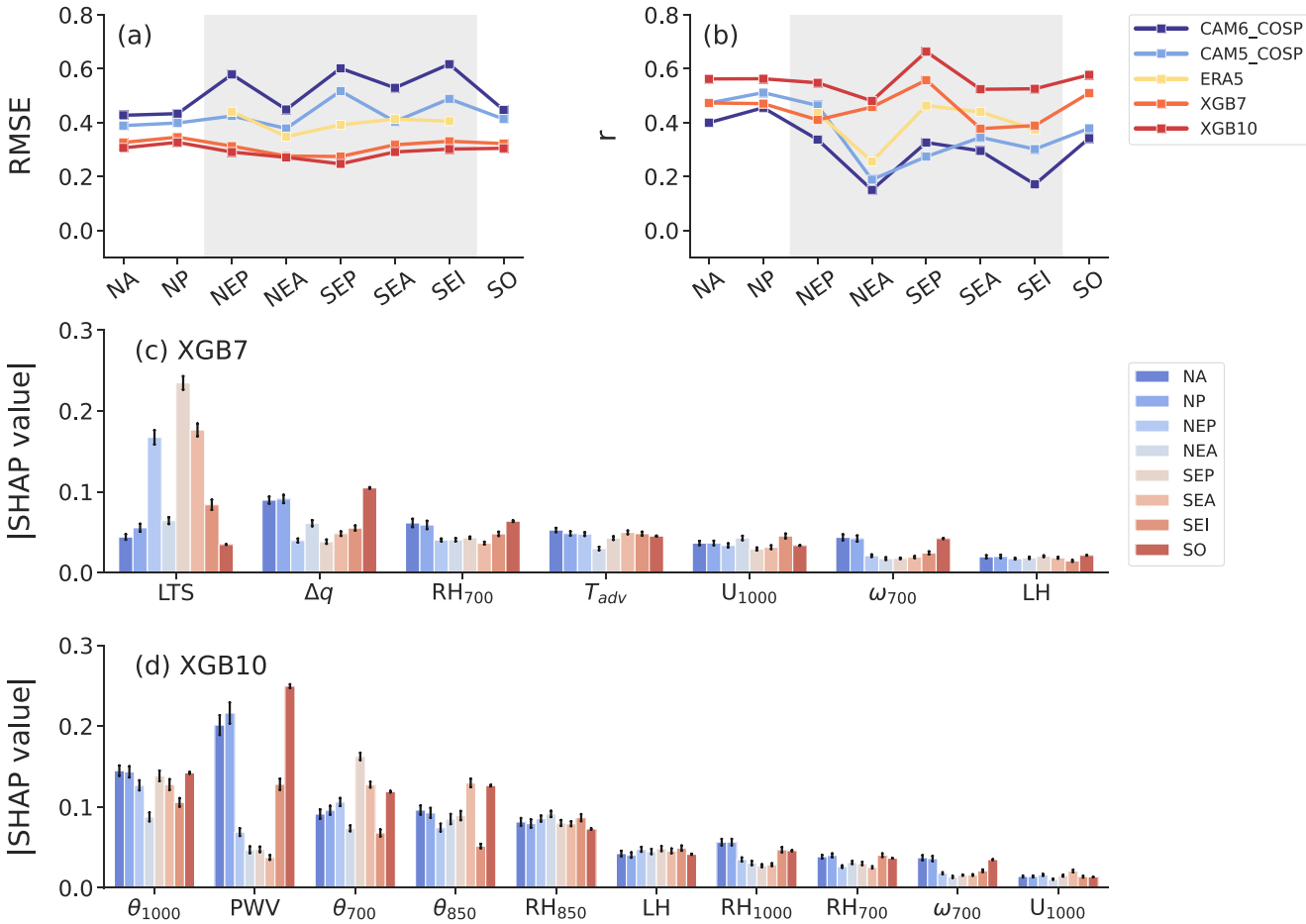

Figure 3. (a) Root mean squared error (RMSE) and (b) correlation coefficients (r) of predicted low-cloud fraction (LCF) from different models among eight selected regions where stratocumulus clouds dominate. All the r values are statistically significant at the 0.05 level. The gray-shaded areas mark the subtropics. Panels (c)–(d) show the absolute SHAP values of each predictor in the eight regions from XGB7 and XGB10, respectively. The predictors in panels (c)–(d) are arranged in descending order based on the region-average absolute SHAP values. The error bars represent the expanded standard error of 50 for visualization.

Northeast Pacific (NEP), Northeast Atlantic (NEA), Southeast Pacific (SEP), Southeast Atlantic (SEA), Southeast Indian Ocean (SEI), and Southern Ocean (SO), identified by Klein and Hartmann (1993). Overall, CAM6 has the highest RMSEs in these regions, underperformed by CAM5 and ERA5 as far as LCF is concerned. Both CAM models display a markedly higher RMSE for LCF in the subtropics (marked by the gray-shaded areas in Figure 3a) compared to the midlatitudes (NA, NP, and SO). In contrast, two ML models exhibit consistently low RMSEs across all eight regions, which have remarkably reduced the biases in the subtropical LCF (especially in NEP, SEP, and SEI), by almost half relative to CAM6. Additionally, XGB7 and XGB10 exhibit similar low RMSEs, suggesting their comparable effectiveness in addressing the underrepresentation of low clouds. However, when evaluating the models' capability to capture spatiotemporal variations in LCF, disparities between XGB7 and XGB10 become evident. XGB7's performance, in terms of r values (Figure 3b), generally falls behind XGB10's, except in SEA. This contrast further underscores the limitations inherent in relying solely on the seven well-known physical processes in Qu et al. (2015) for representing LCF variations. As a result, XGB7's performance (referencing XGB10) degrades to levels comparable with CAM5 in NA and SEA, slightly below CAM5 in NP and NEP.

Through SHAP analysis of ML prediction results, we can elucidate the relative importance of various MFs in LCF predictions. Figures 3c and 3d summarize the SHAP values for each MF in eight selected regions for XGB7 and XGB10, respectively. In both ML models, MFs display pronounced regional dependence. In XGB7, LTS contributes the most in the subtropics, whereas contributions of Δq and RH_{700} dominate in the midlatitudes. In XGB10, θ_{700} is the most important MF in the subtropics, which is similar to LTS in XGB7. But a remarkable

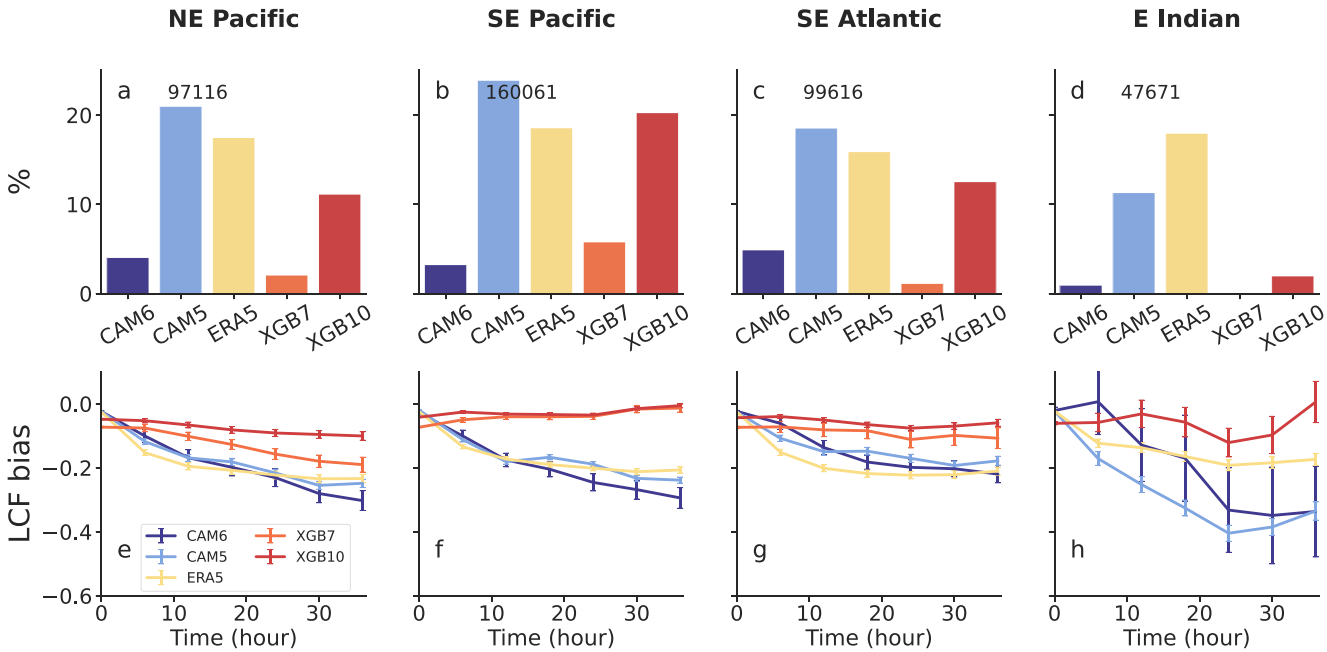

Figure 4. (a–d) Barplot of the percentage of predicted stratocumulus-to-cumulus transition (SCT) occurrence frequency referencing the observed one in the region of Northeast Pacific, Southeast Pacific, Southeast Atlantic, and East Indian regions, respectively. The number shows the sample size of observed SCT cases. (e–h) Time series of ensemble-mean low-cloud fraction (LCF) bias (model minus observation) along all the trajectories in these four regions from different models. Note that XGB7 is removed from panel (h) due to too few SCT cases predicted. The error bars represent the expanded standard error of 5 for visualization.

difference from XGB7 is that PWV dominates in the midlatitude LCF predictions likely because water vapor content is more limited in the midlatitudes (Mao et al., 2017). This explains the degradation of XGB7 in representing LCF variations in these regions due to its lack of considering the effect of moisture supply. It is also worth noting that the SHAP values for PWV are most diverse among the regions, highlighting the regional importance of PWV.

3.3. Evaluation of the SCT Prediction

Figures 4a–4d illustrate the percentage of simulated SCT case counts compared to observed ones in four subtropical regions where SCT predominantly occurs. It is evident that across four regions all models notably underestimate the occurrence of the SCT, capturing less than 25% of observations, with CAM6 and XGB7 exhibiting the poorest performance. The ML models have worse performance compared to the CAM5 and ERA5, even for XGB10. The primary cause for ML models' deficiency in capturing the SCT occurrence is due to their limitation in predicting the extremely high value of LCF (i.e., greater than 0.9), one important criterion used for selecting SCT cases. The issue of predicting extreme values in ML models is commonly seen in other studies, such as the observed smoothing effect when predicting aerosol optical depth fields via ConvLSTM (Wang et al., 2024). Figures 4e–4h present the evolution of ensemble-mean LCF biases (model minus observation) along the selected SCT trajectories. It is usually observed that LCF displays a decreasing trend during the SCT due to enhanced entrainment drying and/or precipitation (Bretherton et al., 1999; Sandu & Stevens, 2011; H. Zhang et al., 2023; Zheng, Zhang, & Li, 2021). However, the CAMs and ERA5 tend to simulate a more rapid decrease in LCF than observations. This is corroborated in Figures 4e–4h, showing that they all have a growing negative bias in LCF over time. Their relative performance shows a regional dependence. For example, ERA5 surpasses the CAMs in East Indian areas but falls short in the Southeast Atlantic, maintaining comparable performance in the Northeast and Southeast Pacific regions. Nevertheless, ML models, especially XGB10, align their predicted LCF more closely with observed values with little dependence on time, outperforming the CAMs and ERA5 across all four regions (particularly in the Southeast Pacific), improving the too-rapid SCT issue. It is noted that non-COSP-simulated LCF in the CAMs is used for evaluation in the SCT as employing COSP for MODIS will lead to non-global results per each output and interrupt the LCF tracking along the trajectories. Since we focus on the SCT areas where high-level clouds are few, the influence of employing COSP on our conclusion is minor.

4. Summary

Stratocumulus clouds over cold oceans are frequently underestimated by models, which also transition to cumulus clouds too quickly, due to inadequate parameterization of cloud processes. This study exploits the credentials of machine learning (ML) models (specifically XGBoost) in predicting the low-cloud fraction (LCF) with respect to several traditional cloud schemes employed in two generations of community atmospheric models (CAM5 and CAM6) and ERA5 reanalysis data. To mitigate the impact of simulation errors related to large-scale meteorological factors, we nudged the wind speeds, temperature, and moisture toward ERA5 in CAM5 and CAM6. Our findings show that ML models significantly improve LCF predictions in terms of both climatological cloud distribution patterns, and some error metrics such as root mean squared errors and correlation coefficients with reference to the satellite LCF product of the CERES SYN Ed4. This improvement is notable across cloud and dynamics regimes, highlighting the ML models' independence from the conditions that heavily influence LCF predictions in global climate models (GCMs).

Employing an interpretable ML approach (SHAP), we did an explainability analysis, and found that including the effect of moisture source in ML models is crucial to representing spatiotemporal variations in LCF in the mid-latitudes, noting that this key variable has not been used in popular cloud fraction parametrizations. Furthermore, through a Lagrangian framework, ML models are found to excel in capturing the LCF variations during the stratocumulus-to-cumulus transition (SCT). Overall, ML models demonstrate great potential to address the longstanding issues of "too few" low clouds and "too rapid" SCT in GCMs. However, the tree-based ML models used in our study exhibit some limitations: (a) explaining less than 50% of observed LCF variances; (b) substantially underestimating the occurrence of SCT cases. These challenges highlight the need for further research to either improve or adopt new ML models before their integration as LCF parameterization schemes in climate modeling.

Data Availability Statement

CAM5 and CAM6 are run within the host model, the Community Earth System Model 2 (CESM2), and its code is available at Danabasoglu et al. (2020). The scripts for setting up CAM5 and CAM6, as well as for optimizing and training ML models (XGB10 and XGB7), are available at H. Zhang et al. (2024b). The ERA5 reanalysis data are available at Hersbach et al. (2020). The CERES SYN Ed4 products are available at CERES Science Team (2021).

Acknowledgments
This study is supported by the Department of Energy (DOE) Atmospheric System Research program (DESC0022919) and the National Science Foundation (AGS2126098). YZ is supported by the DOE Early Career Grant (DE-SC0024185). We acknowledge high-performance computing support from Derecho: HPE Cray EX System (https://doi.org/10.5065/qx9a-pg09) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. We thank Ryan Eastman for providing MATLAB codes for trajectory calculations. We thank two anonymous reviewers for their constructive comments, which greatly improved the quality of the manuscript.

References

Albrecht, B. A., Bretherton, C., Johnson, D., Scubert, W. H., & Frisch, A. S. (1995). The Atlantic Stratocumulus Transition Experiment—ASTEX. *Bulletin of the American Meteorological Society*, *76*(6), 889–904. https://doi.org/10.1175/1520-0477(1995)076<0889:TASTE>2.0.CO;2

Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J. L., Klein, S. A., et al. (2011). COSP: Satellite simulation software for model assessment. *Bulletin of the American Meteorological Society*, *92*(8), 1023–1043. https://doi.org/10.1175/2011BAMS2856.1

Bogenschutz, P. A., Gettelman, A., Morrison, H., Larson, V. E., Craig, C., & Schanen, D. P. (2013). Higher-order turbulence closure and its impact on climate simulations in the community atmosphere model. *Journal of Climate*, *26*(23), 9655–9676. https://doi.org/10.1175/JCLI-D-13-00075.1

Bony, S., & Dufresne, J. L. (2005). Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophysical Research Letters*, *32*(20), L20806. https://doi.org/10.1029/2005GL023851

Bretherton, C. (2015). Insights into low-latitude cloud feedbacks from high-resolution models. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, *373*(2054), 20140415. https://doi.org/10.1098/rsta.2014.0415

Bretherton, C., & Blossey, P. N. (2014). Low cloud reduction in a greenhouse-warmed climate: Results from Lagrangian les of a subtropical marine cloudiness transition. *Journal of Advances in Modeling Earth Systems*, *6*(1), 91–114. https://doi.org/10.1002/2013MS000250

Bretherton, C., Blossey, P. N., & 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. https://doi.org/10.1002/jame.20019

Bretherton, C., Krueger, S. K., Wyant, M. C., Bechtold, P., Stevens, B., & Teixeira, J. (1999). A GCSS boundary-layer cloud model intercomparison study of the first astex Lagrangian experiment. *Boundary-Layer Meteorology*, *93*(3), 341–380. https://doi.org/10.1023/a:1002005429969

Ceppi, P., Brient, F., Zelinka, M. D., & Hartmann, D. L. (2017). Cloud feedback mechanisms and their representation in global climate models. *WIREs Climate Change*, *8*(4), e465. https://doi.org/10.1002/wcc.465

CERES Science Team. (2021). The CERES SYN1deg Ed4A products. [Dataset]. *NASA Atmospheric Science Data Center (ASDC)*. Retrieved from https://ceres-tool.larc.nasa.gov/ord-tool/jsp/SYN1degEd41Selection.jsp

Chen, G., Wang, W., Bao, Q., & Li, J. (2022). Evaluation of simulated cloud diurnal variation in CMIP6 climate models. *Journal of Geophysical Research: Atmospheres*, *127*(6), e2021JD036422. https://doi.org/10.1029/2021JD036422

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Chen, Y., Yamaguchi, T., Bogenschutz, P. A., & Feingold, G. (2021). Model evaluation and intercomparison of marine warm low cloud fractions with neural network ensembles. *Journal of Advances in Modeling Earth Systems*, *13*(11), e2021MS002625. https://doi.org/10.1029/2021MS002625

Chen, Z., Wang, M., Zhang, H., Lin, S., Guo, Z., Jiang, Y., & Zhou, C. (2022). Long-term change in low-cloud cover in Southeast China during cold seasons. *Atmospheric and Oceanic Science Letters*, *15*(6), 100222. https://doi.org/10.1016/j.aosl.2022.100222

Christensen, M. W., Ma, P. L., Wu, P., Varble, A. C., Mülmenstädt, J., & Fast, J. D. (2023). Evaluation of aerosol-cloud interactions in E3SM using a Lagrangian framework. *Atmospheric Chemistry and Physics*, *23*(4), 2789–2812. https://doi.org/10.5194/acp-23-2789-2023

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., et al. (2020). The Community Earth System Model Version 2 (CESM2) [Software]. *NCAR*. Retrieved from https://www.cesm.ucar.edu/models/cesm2

Doelling, D. R., Loeb, N. G., Keyes, D. F., Nordeen, M. L., Morstad, D., Nguyen, C., et al. (2013). Geostationary enhanced temporal interpolation for ceres flux products. *Journal of Atmospheric and Oceanic Technology*, *30*(6), 1072–1090. https://doi.org/10.1175/JTECH-D-12-00136.1

Donner, L. J., Wyman, B. L., Hemler, R. S., Horowitz, L. W., Ming, Y., Zhao, M., et al. (2011). The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3. *Journal of Climate*, *24*(13), 3484–3519. https://doi.org/10.1175/2011JCLI3955.1

Eastman, R., Terai, C. R., Grosvenor, D. P., & Wood, R. (2021). Evaluating the Lagrangian evolution of subtropical low clouds in GCMs using observations: Mean evolution, time scales, and responses to predictors. *Journal of the Atmospheric Sciences*, *78*(2), 553–572. https://doi.org/10.1175/JAS-D-20-0178.1

Eastman, R., & Wood, R. (2018). The competing effects of stability and humidity on subtropical stratocumulus entrainment and cloud evolution from a Lagrangian perspective. *Journal of the Atmospheric Sciences*, *75*(8), 2563–2578. https://doi.org/10.1175/JAS-D-18-0030.1

ECMWF. (2016). IFS Documentation CY41R2 - Part IV: Physical Processes. https://doi.org/10.4324/9781315069258-11

Edwards, P. N. (2011). History of climate modeling. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(1), 128–139. https://doi.org/10.1002/wcc.95

Golaz, J.-C., Larson, V. E. E., & Cotton, W. R. R. (2002a). A PDF-based model for boundary layer clouds. Part I: Method and model description. *Journal of the Atmospheric Sciences*, *59*(24), 3540–3551. https://doi.org/10.1175/1520-0469(2002)059<3540:APBMFB>2.0.CO;2

Golaz, J.-C., Larson, V. E. E., & Cotton, W. R. R. (2002b). A PDF-based model for boundary layer clouds. Part II: Model results. *Journal of the Atmospheric Sciences*, *59*(24), 3552–3571. https://doi.org/10.1175/1520-0469(2002)059<3552:APBMFB>2.0.CO;2

Hersbach, H., Bell, B., Berrisford, P., Zhang, M., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. [Dataset]. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. https://doi.org/10.24381/cds.bd0915c6

Houze, R. A., Jr. (2014). *Cloud dynamics* (2nd ed.). Academic Press. Retrieved from https://shop.elsevier.com/books/cloud-dynamics/houze-jr/978-0-12-374266-7

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., et al. (2013). The community earth system model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, *94*(9), 1339–1360. https://doi.org/10.1175/BAMS-D-12-00121.1

Jian, B., Li, J., Wang, G., Zhao, Y., Li, Y., Wang, J., et al. (2021). Evaluation of the CMIP6 marine subtropical stratocumulus cloud albedo and its controlling factors. *Atmospheric Chemistry and Physics*, *21*(12), 9809–9828. https://doi.org/10.5194/acp-21-9809-2021

Klein, S. A., & Hartmann, D. L. (1993). The seasonal cycle of low stratiform clouds. *Journal of Climate*, *6*(8), 1587–1606. https://doi.org/10.1175/1520-0442(1993)006<1587:TSCOLS>2.0.CO;2

Larson, V. E. (2017). CLUBB-SILHS: A parameterization of subgrid variability in the atmosphere. Retrieved from http://arxiv.org/abs/1711.03675

Larson, V. E., & Golaz, J.-C. (2005). Using probability density functions to derive consistent closure relationships among higher-order moments. *Monthly Weather Review*, *133*(4), 1023–1042. https://doi.org/10.1175/MWR2902.1

Larson, V. E., Golaz, J.-C., & Cotton, W. R. (2002). Small-scale and mesoscale variability in cloudy boundary layers: Joint probability density functions. *Journal of the Atmospheric Sciences*, *59*(24), 3519–3539. https://doi.org/10.1175/1520-0469(2002)059<3519:SSAMVI>2.0.CO;2

Larson, V. E., Schanen, D. P., Wang, M., Ovchinnikov, M., & Ghan, S. (2012). PDF parameterization of boundary layer clouds in models with horizontal grid spacings from 2 to 16 km. *Monthly Weather Review*, *140*(1), 285–306. https://doi.org/10.1175/MWR-D-10-05059.1

Li, J., An, X., Li, Q., Wang, C., Yu, H., Zhou, X., & Geng, Y. (2022). Application of XGBoost algorithm in the optimization of pollutant concentration. *Atmospheric Research*, *276*, 106238. https://doi.org/10.1016/j.atmosres.2022.106238

Li, T., Wang, M., Guo, Z., Yang, B., Xu, Y., Han, X., & Sun, J. (2022). An updated CLUBB PDF closure scheme to improve low cloud simulation in CAM6. *Journal of Advances in Modeling Earth Systems*, *14*(12), 1–21. https://doi.org/10.1029/2022MS003127

Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2018). Consistent individualized feature attribution for tree ensembles. Retrieved from http://arxiv.org/abs/1802.03888

Lundberg, S. M., & Lee, S.-I. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 2, pp. 4766–4775). Retrieved from http://arxiv.org/abs/1705.07874

Mao, K., Chen, J., Li, Z., Ma, Y., Song, Y., Tan, X., & Yang, K. (2017). Global water vapor content decreases from 2003 to 2012: An analysis based on MODIS data. *Chinese Geographical Science*, *27*(1), 1–7. https://doi.org/10.1007/s11769-017-0841-6

Medeiros, B., & Stevens, B. (2011). Revealing differences in GCM representations of low clouds. *Climate Dynamics*, *36*(1), 385–399. https://doi.org/10.1007/s00382-009-0694-5

Nam, C., Bony, S., Dufresne, J. L., & Chepfer, H. (2012). The too few, too bright tropical low-cloud problem in CMIP5 models. *Geophysical Research Letters*, *39*(21), L21801. https://doi.org/10.1029/2012GL053421

Neale, R. B., Gettelman, A., Park, S., Chen, C.-C., Lauritzen, P. H., Williamson, D. L., et al. (2012). Description of the NCAR Community Atmosphere Model (CAM 5.0). Retrieved from https://www2.cesm.ucar.edu/models/cesm1.0/cam/docs/description/cam5_desc.pdf

Neale, R. B., Richter, J. H., Conley, A. J., Park, S., Lauritzen, P. H., Gettelman, A., et al. (2010). Description of the NCAR Community Atmosphere Model (CAM 4.0). Retrieved from https://www2.cesm.ucar.edu/models/ccsm4.0/cam/docs/description/cam4_desc.pdf

Pan, B., Anderson, G. J., Goncalves, A., Lucas, D. D., Bonfils, C. J. W., & Lee, J. (2022). Improving seasonal forecast using probabilistic deep learning. *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002766. https://doi.org/10.1029/2021MS002766

Qu, X., Hall, A., Klein, S. A., & Deangelis, A. M. (2015). Positive tropical marine low-cloud cover feedback inferred from cloud-controlling factors. *Geophysical Research Letters*, *42*(18), 7767–7775. https://doi.org/10.1002/2015GL065627

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Sandu, I., & Stevens, B. (2011). On the factors modulating the stratocumulus to cumulus transitions. *Journal of the Atmospheric Sciences*, *68*(9), 1865–1881. https://doi.org/10.1175/2011JAS3614.1

Schuddeboom, A. J., & McDonald, A. J. (2021). The southern ocean radiative bias, cloud compensating errors, and equilibrium climate sensitivity in CMIP6 models. *Journal of Geophysical Research: Atmospheres*, *126*(22), e2021JD035310. https://doi.org/10.1029/2021JD035310

Sellevold, R., & Vizcaino, M. (2021). First application of artificial neural networks to estimate 21st century Greenland ice sheet surface melt. *Geophysical Research Letters*, *48*(16), e2021GL092449. https://doi.org/10.1029/2021GL092449

Soden, B. J., & Vecchi, G. A. (2011). The vertical distribution of cloud feedback in coupled ocean-atmosphere models. *Geophysical Research Letters*, *38*(12), L12704. https://doi.org/10.1029/2011GL047632

Tiedtke, M. (1993). Representation of clouds in large-scale models. *Monthly Weather Review*, *121*(11), 3040–3061. https://doi.org/10.1175/1520-0493(1993)121<3040:ROCILS>2.0.CO;2

Vial, J., Dufresne, J.-L. L., & Bony, S. (2013). On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates. *Climate Dynamics*, *41*(11–12), 3339–3362. https://doi.org/10.1007/s00382-013-1725-9

Vignesh, P. P., Jiang, J. H., Kishore, P., Su, H., Smay, T., Brighton, N., & Velicogna, I. (2020). Assessment of CMIP6 cloud fraction and comparison with satellite observations. *Earth and Space Science*, *7*(2), 1–21. https://doi.org/10.1029/2019EA000975

Wang, Y., Schoeberl, M. R., Esmaili, R. B., & Liu, J. (2024). Gap-filling AOD data using deep learning techniques in satellite imagery. In *104th American Meterology Society Meeting, Baltimore*. Retrieved from https://ams.confex.com/ams/104ANNUAL/meetingapp.cgi/Paper/437382

Wood, R. (2012). Stratocumulus clouds. *Monthly Weather Review*, *140*(8), 2373–2423. https://doi.org/10.1175/MWR-D-11-00121.1

Wood, R., & Bretherton, C. (2006). On the relationship between stratiform low cloud cover and lower-tropospheric stability. *Journal of Climate*, *19*(24), 6425–6432. https://doi.org/10.1175/JCLI3988.1

Wyant, M. C., Bretherton, C., Rand, H. A., & Stevens, D. E. (1997). Numerical simulations and a conceptual model of the stratocumulus to trade cumulus transition. *Journal of the Atmospheric Sciences*, *54*(1), 168–192. https://doi.org/10.1175/1520-0469(1997)054<0168:NSAACM>2.0.CO;2

Xu, K.-M., & Randall, D. A. (1996). A semiempirical cloudiness parameterization for use in climate models. *Journal of the Atmospheric Sciences*, *53*(21), 3084–3102. https://doi.org/10.1175/1520-0469(1996)053<3084:ASCPFU>2.0.CO;2

Zhang, H., Wang, M., Guo, Z., Zhou, C., Zhou, T., Qian, Y., et al. (2018). Low-cloud feedback in CAM5-CLUBB: Physical mechanisms and parameter sensitivity analysis. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2844–2864. https://doi.org/10.1029/2018MS001423

Zhang, H., Zheng, Y., Lee, S. S., & Li, Z. (2023). Surface-atmosphere decoupling prolongs cloud lifetime under warm advection due to reduced entrainment drying. *Geophysical Research Letters*, *50*(10), e2022GL101663. https://doi.org/10.1029/2022GL101663

Zhang, H., Zheng, Y., & Li, Z. (2024a). Evaluation of stratocumulus evolution under contrasting temperature advections in CESM2 through a Lagrangian framework. *Geophysical Research Letters*, *51*(4), e2023GL106856. https://doi.org/10.1029/2023GL106856

Zhang, H., Zheng, Y., & Li, Z. (2024b). Scripts for Paper Titled "Improving low-cloud fraction prediction through machine learning" (v2.0) [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.12002758

Zhang, M., Bretherton, C. S., Blossey, P. N., Austin, P. H., Bacmeister, J. T., Bony, S., et al. (2013). CGILS: Results from the first phase of an international project to understand the physical mechanisms of low cloud feedbacks in single column models. *Journal of Advances in Modeling Earth Systems*, *5*(4), 826–842. https://doi.org/10.1002/2013MS000246

Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., et al. (2018). The GFDL Global Atmosphere and Land Model AM4.0/LM4.0: 1. Simulation characteristics with prescribed SSTs. *Journal of Advances in Modeling Earth Systems*, *10*(3), 691–734. https://doi.org/10.1002/2017MS001208

Zheng, Y., Zhang, H., & Li, Z. (2021). Role of surface latent heat flux in shallow cloud transitions: A mechanism-denial LES study. *Journal of the Atmospheric Sciences*, *78*(9), 2709–2723. https://doi.org/10.1175/JAS-D-20-0381.1

Zheng, Y., Zhang, H., Rosenfeld, D., Lee, S.-S. S., Su, T., & Li, Z. (2021). Idealized large-eddy simulations of stratocumulus advecting over cold water. Part I: Boundary layer decoupling. *Journal of the Atmospheric Sciences*, *78*(12), 4089–4102. https://doi.org/10.1175/JAS-D-21-0108.1

References From the Supporting Information

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Snoek, J., Larochelle, H., Adams, R. P., & Jeffery, C. (2012). Practical Bayesian optimization of machine learning algorithms. *Religion and the Arts*, *17*(1–2), 57–73. https://doi.org/10.48550/ARXIV.1206.2944