https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

Geoscientific Model Development
Discussions
Open Access
EGU

Deep Learning Driven Simulations of Boundary Layer Cloud over the US Southern Great Plains

Tianning Su[1*], Yunyan Zhang[1]

[1]Lawrence Livermore National Laboratory, Livermore, CA, USA

*Corresponding authors: su10@llnl.gov

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

Geoscientific Model Development Discussions
Open Access

Abstract. This study developed a deep learning model to simulate the complex dynamics of boundary layer clouds (BLCs) over the US Southern Great Plains. Using over twenty years of extensive observations from the Atmospheric Radiation Measurement program for training and validation, the model diagnoses the BLCs from the perspective of cloud-land coupling. Morning meteorological profiles set as the initial conditions and then identifying triggers for BLCs formation from surface meteorology. The deep learning model offer accurate simulation of the convection initiation and cloud base of BLCs. In comparison with reanalysis data (i.e., ERA-5 and MERRA-2), it provides a notable improvement in the vertical structure of low clouds from a climatological perspective. The deep learning model can serve as the cloud parameterization and extend to analyzing stratiform and cumulus clouds within reanalysis frameworks, offering insights into improving the simulation of BLCs. By quantifying biases due to various meteorological factors and parameterizations, this deep learning-driven approach bridges the observational-modeling divide. Surface humidity and parameterization emerge as key limiting factors to affect the representation of BLCs in the reanalysis data. This deep learning approach holds promise for improving the convection parameterization and advancing model diagnostics in weather forecasting and climate modelling.

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

1 Introduction

Boundary layer clouds (BLCs), comprising primarily of stratiforms and shallow cumuli, exert a profound influence on the Earth's radiative balance and climate system (Betts, 2009; Teixeira and Hogan, 2002; Lu et al., 2013; Golaz et al., 2002). Their formation and evolution within the planetary boundary layer (PBL) are critically shaped by the interactions between surface processes and atmospheric dynamics (Miao et al., 2019; Berg and Kassianov, 2008; Zhang and Klein, 2013; Guo et al., 2019; Zhang et al., 2017). These clouds, which frequently form in the PBL's entrainment zone, are the critical part for weather prediction and climate modeling, holding the key to understanding land-atmosphere interactions (Caldwell et al., 2021; Bretherton et al., 2007; Wang et al., 2020, 2023; Moeng et al., 1996; Su et al. 2023; Zhang and Klein, 2010; Guo et al., 2019).

Numerous studies have been dedicated to investigating the dynamics of boundary layer clouds, highlighting the pivotal role of land surface in modulating cloud formation and affecting the spatial and temporal distribution of low clouds (Rieck et al., 2014; Xiao et al., 2018; Lee et al., 2019; Fast et al., 2019b; Tang et al., 2018; Tang et al., 2019; Tao et al., 2019; Tian et al., 2022; Qian et al., 2023). Despite considerable advancements in observations and modeling capabilities, simulating the boundary layer clouds remains a significant challenge, largely due to the complex feedback mechanisms between land surface fluxes, PBL processes, and cloud microphysics (Miao et al., 2019; Lareau et al., 2018; Lu et al., 2011; Fast et al., 2019a; Wang et al., 2014; Yang et al., 2018; Nogherotto et al., 2016). These challenges are compounded when attempting to

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

represent such processes in global and regional climate models, where the fine-scale interactions are often parameterized in a coarse-resolution grid due to computational constraints (Bretherton et al., 2007; Moeng et al., 1996). In addition, different cloud regimes exhibit varied cloud-land interactions that complicate their representation and present challenges for observational studies and modeling efforts (Tang et al., 2018; Qian et al., 2023; Sakaguchi et al., 2022; Poll et al., 2022; Tao et al., 2021).

As an emerging tool, machine learning (ML) has been widely employed for a variety of environmental and atmospheric studies (e.g., McGovern et al., 2017; Gagne et al., 2019; Vassallo et al., 2020; Cadeddu et al., 2009; Molero et al., 2022; Guo et al., 2024). Specifically, ML techniques are increasingly being employed to simulate and estimate convection and precipitation, which are crucial for accurate weather forecasting and climate modeling (Mooers et al., 2021; Wang et al., 2020; O'Gorman et al., 2018; Gentine et al., 2018; Zhang et al., 2021). For example, Rasp (2020) presents algorithms for the implementation of coupled learning in cloud-resolving models and the super parameterization framework. Similarly, ML tools have been applied to leverage observational data for the refinement of convection parameterizations, offering more insights into convective dynamics (O'Gorman et al., 2018; Gentine et al., 2018; Zhang et al., 2021). Moreover, Haynes et al. (2022) develop pixel-based ML-based methods of detecting low clouds, with a focus on improving detection in multilayer cloud situations and specific attention given to improving the cloud characterization.

Despite the considerable advancements brought by ML, there are persistent challenges in accurately simulating the vertical structure of clouds, as well as their

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

complex relationships with land surface. To address these complexities, this study developed an advanced deep-learning framework to simulate the BLCs, using comprehensive data from the Atmospheric Radiation Measurement (ARM) program at the Southern Great Plains (SGP) site. This framework is designed to simulate the triggers, progression, and structural structure of boundary layer clouds, placing a particular emphasis on cloud-land coupling mechanisms. By assimilating morning radiosonde observations with diurnal-varying surface fluxes and meteorological data, this deep learning model is uniquely positioned to unravel the complex initiation and evolution of low clouds, especially those coupled with land surface processes.

Furthermore, the critical assessment of our model in comparison with existing reanalysis datasets, including MERRA-2 and ERA-5, highlights the improvement in representing cloud vertical structure. Our study analyzed the model's performance across various cloud regimes, such as stratiform and cumulus. By serving as the cloud parameterization in the reanalysis data, this model advanced the capability of low cloud simulations within reanalysis frameworks. By undertaking this endeavor, we strive to narrow the existing gaps in boundary layer clouds between field observations and modeling, thereby enriching our understanding of the convective processes.

2 Data and instruments

2.1 Comprehensive observations at the US Southern Great Plains

The ARM program, funded by the U.S. Department of Energy, plays a pivotal role at the SGP site in Oklahoma (36.607°N, 97.488°W). Recognized globally as a leading

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.
Geoscientific Model Development Discussions

climate research facility, the ARM SGP site has been collecting a wealth of meteorological and radiative measurements, offering data that spans over two decades. This study utilized these extensive field observations during 1998-2020 to explore the simultions of BLCs from the deep learning techniques.

Our study employs the Active Remote Sensing of Clouds (ARSCL) product, which integrates lidars, ceilometer, and cloud radar data to define cloud boundaries, utilizing the best estimates from lidar for the lowest cloud bases (Clothiaux et al. 2000, 2001; Kollias et al. 2020). Based on the comprehensive information of cloud vertical structure and temporal evolution from the ARSCL dataset, Xie et al. (2010) offers detailed cloud fraction profiles at the hourly resolution in ARM BEST ESTIMATE DATA PRODUCTS (ARMBE).

Central to our analysis are the comprehensive thermodynamic profiles obtained from radiosondes (SONDE). Launched routinely at multiple times daily, these SONDE measurements offer detailed information into the thermodynamic state of the PBL and the free atmosphere. The operation and technical aspects of the ARM SONDE are detailed in Holdridge et al. (2011).

Meanwhile, We use a variety of meteorological parameters (i.e., temperature, relative humidity, wind, and pressure) at the surface level from collocated surface meteorology systems (MET) at the ARM SGP site. These surface meteorological parameters play roles in the formation and development of BLCs. Furthermore, the site also provides data on surface sensible and latent heat fluxes. An ARM value-added product called the best-estimate fluxes from energy balance Bowen ratio measurements

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

and Bulk Aerodynamic calculations (BAEBBR), was generated to replace the energy balance Bowen ratio flux measurements with a bulk aerodynamic estimation when the Bowen Ratio has a range of -1.6 and -0.45 (Cook, 2018). The replacements generally happen for the measurements at a 30-min temporal resolution under the low sensible heat scenario. We use the BAEBBR data along with the surface meteorology as the key input for the deep learning model.

2.2 Boundary layer clouds

BLCs demonstrate complex evolutions and interact with boundary layer and land surface processes. Tracking the initiation, development, and lifecycle of BLCs is crucial for understanding the convection in our climate system. We treat BLCs as synonymous with land-coupled clouds, in contrast to those clouds are decoupled from PBL and land surface. In this regard, Su et al. (2022) devised a lidar-based method to discern cloud-land coupling, leveraging the vertical coherence and temporal continuity of the PBL. This approach, combined with cloud base height measurements from ceilometers and surface-based Lifting Condensation Level (LCL) calculations as proposed by Romps (2017), forms the foundation for identifying coupled low-level clouds in our study. The methodology for determining PBLH, as outlined by Su et al. (2020), established a long-term record of PBLH at the SGP. The resulting data are publicly available through the ARM database (https://www.arm.gov/data/data-sources/pblht-206). Coupled clouds are identified when the cloud base height (CBH), as derived from ceilometer, aligns with the lidar-detected PBL top and the calculated

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

LCL within a certain range (Su et al. 2022). This alignment is indicative of clouds that are directly influenced by surface-driven processes. Meanwhile, we apply a cloud thickness threshold of less than 4 km to analyze BLCs.

Within the scope of land-coupled clouds, we further classify observed BLCs into cumulus and stratiform categories. For cumulus clouds, two criteria are applied: (1) cloud formations emerge after sunrise, ensuring that they are driven by local convective processes, and (2) there is an absence of overcasting clouds. Conversely, stratiform clouds are identified by prolonged overcasting conditions during the daytime, typically lasting more than three hours, with cloud fractions exceeding 90% as per ARSCL data. Based on the criteria, we identified 940 days categorized under the cumulus regime, distributed as 21%, 56%, 17%, and 6% across Spring, Summer, Fall, and Winter, respectively. Similarly, we identified 657 days falling within the stratiform clouds regime, with respective seasonal distributions of 37%, 12%, 23%, and 28%. This classification of cloud types to filter the BLCs based on the concept of cloud-land coupling is important for the training and analysis of the deep learning model. We further attempt to use the comprehensive cloud observations at the SGP to build the deep learning model.

2.3 Reanalysis data

In this study, we also use reanalysis datasets from the European Centre for Medium-Range Weather Forecasts' fifth-generation global reanalysis (ERA-5, Hersbach et al., 2020) and NASA's Modern-Era Retrospective analysis for Research

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

Geoscientific Model Development Discussions

and Applications Version 2 (MERRA-2, Gelaro et al., 2017).

As the state-of-art reanalysis data, the ERA-5 is produced by the Integrated Forecasting System (IFS) and a data assimilation system at a fine spatial resolution of 0.25° x 0.25°. ERA-5 reanalysis data are obtained from the Copernicus Climate Data Store. This dataset employs a prognostic cloud scheme capable of capturing the evolution of cloud dynamics over consecutive time steps (Tiedtke 1993), a feature that enhances its utility in time-dependent climate studies. We also use ERA-5 data to obtain the cloud and atmospheric information, which provides the hourly measurements at a 0.25° - 0.25° longitude-latitude grid. The vertical resolution of ERA-5 data is 25 hPa in the lower atmosphere (700 – 1000 hPa).

The MERRA-2 reanalysis data use a new version of the Goddard Earth Observing System Data Assimilation System Version 5, which is a advanced system coupling a global atmospheric general circulation model to NCEP's Grid-point Statistical Interpolation analysis (Randles et al., 2017). The MERRA-2 reanalysis has a spatial resolution of $2/3^{\circ} \times 1/2^{\circ}$ (longitude – latitude). MERRA-2 utilizes a diagnostic cloud scheme, focusing on the immediate state of clouds. This dataset specializes in the representation of the hydrological cycle and cloud information, which are widely used in multiple studies (e.g., Yeo et al., 2022; Kuma, 2020; Miao et al., 2019). MERRA-2's reanalysis provides detailed hourly low cloud fraction data and tri-hourly vertical cloud fraction profiles.

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

3 Deep Learning Framework for Simulating Boundary Layer Clouds

3.1 Integrated Deep Learning Models for Cloud Simulation

This study developed an integrated deep learning approach to simulate BLC. These models are purpose-built to simulate the initiation, positioning, and vertical extent of BLCs. Figure 1 demonstrates the design of our deep learning framework. The core of this research integrates three distinct deep learning models, each responsible for a critical aspect of cloud simulation: the triggering of cloud formation, the determination of cloud position, and the vertical profile of coverage fraction within cloud boundaries, which jointly yielding the comprehensive features of BLCs through multiple hidden layers. Initially, the triggering model evaluates whether cloud formation is likely, producing a value between 0 and 1, with values above 0.5 indicating the presence of clouds. This triggering information then feeds into the other two models: one for determining the cloud's position and the other for calculating the cloud fraction. While the cloud position and cloud fraction models are independent components, they collaborate to depict the vertical cloud fraction profile.

Morning meteorological profiles set as the initial conditions and then identifying triggers for BLC formation from surface meteorology. Each morning profiles have 46 elements from 0-6km. For surface meteorology and fluxes, the inputs include the data at the current hour and the previous hour. The RH profiles and PBL top are highlighted for their significance in boundary layer development. To represent the vertical structure of BLC, we equally segmented the cloud layer from the base to the top into ten levels.

For each of these levels, our deep learning models calculate individual cloud fraction values. These values are then interpolated to create a continuous vertical profile of cloud fraction within the BLC, offering a detailed depiction of the cloud's vertical extent. This model used the TensorFlow Package, developed by Google (https://www.tensorflow.org/).

The deep neural network (DNN) architecture was designed (brown boxes in Figure 1), beginning with an input layer reflective of the selected feature set. The detailed structures for the three models can be found in Table 1. Normalization is a preprocessing technique that often leads to improvements in model training by scaling the input features and target values to a standard range (Raju et al. 2020). We applied the normalization process to both the input and target data to ensure that they have a zero mean and a standard deviation of one. This standardization scales the data to a common range, allowing for a more stable and efficient training process. Subsequent hidden layers were integrated, each imbued with L2 regularization to mitigate overfitting by penalizing complexity.

Batch normalization was implemented at each layer to normalize inputs, ensuring consistent data distribution throughout the network, thereby stabilizing the learning process. A dropout rate of 0.2 further prevented overfitting by randomly omitting neuron connections during training, encouraging the network to learn more robust features. Training was refined with early stopping, ceasing further epochs when the validation loss ceased to improve, and learning rate reduction, systematically

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

decreasing the learning rate upon encountering plateaus in performance improvement. These callbacks were instrumental in honing the model's performance, ensuring convergence to the accurate representation of the BLC. Neuron biases are included in the network's architecture and systematically inserted in the hidden layers (Battaglia et al. 2018).

3.2 Model Training Process and Examples

The construction of our deep learning model suite commenced with the segregation of the comprehensive dataset, sourced from the rich datasets of ARM program, into a training subset (70%) and a validation subset (30%) during 1998-2016. Additionally, we incorporate datasets from 2017-2020 as part of our validation process, specifically focusing on data from the untrained period to assess the model's performance. The training and validations are both using the more than 20-year BLC observations, as well as the ARMBE products. The base of BLC is determined at the lowest altitude where cloud fraction exceeds 1%, and the cloud top is identified at the point where cloud fraction transitions from exceeding 1% to falling below this threshold. In multi-layer systems, the DNN model is trained based on the lowest cloud layer when it is coupled with the land surface. Upon training completion, the model was evaluated, with its performance metrics examined for accuracy and reliability. This methodical and data-driven process balanced complexity with precision, culminating in a robust model capable of simulating BLC features.

These models operate synergistically, with the predicted cloud trigger extending into the models for cloud position and vertical structure (i.e., cloud fraction and cloud thickness). As the example of the model output, Figure 2 offers a comparative display of diurnal cloud fraction profiles over the SGP, contrasting observed data with deep learning simulations. The model accurately simulates the cloud occurrence and cloud base for these cases, aligning well with observations. However, it falls short in simulating cloud top position, especially for stratiform clouds, overestimating cloud tops. It also underestimates maximum cloud fractions for the stratiform clouds. The maximum cloud fraction for stratiform is close to 1, indicating complete coverage, an aspect not fully replicated by the model. The third case also falls into the category of stratiform clouds, characterized by a cloud fraction exceeding 0.9. However, the presence of multiple local maxima within the cloud fraction profile indicates a relatively complex structure. This complexity poses a challenge to the model, as the DNN is not fully capable of capturing the internal variations within the convective system. Instead, the model tends to produce a more uniform cloud fraction across this convective system. Despite these variances, the model-derived cloud bases and occurrence demonstrates high consistency with observations, highlighting its value in the cloud simulations.

3.3 Calculations of Feature Importance and Performance Metric

To elucidate the significance of each variable within our deep learning models, we implemented a permutation importance analysis. This robust, model-agnostic technique assesses each feature's influence on the model's predictive accuracy, which is crucial

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

for assessing DNN (Date and Kikuchi, 2018; Altmann et al. 2010). In this study, the permutation importance method differs slightly based on whether the model's task is regression (cloud position and fraction) or classification (trigger).

For the models predicting cloud position and cloud fraction, which are regression tasks, the Mean Absolute Error (MAE) serves as the performance metric. First, we perform a test run to establish a baseline performance by calculating the MAE of the model using the original, unperturbed validation datasets, which comprises morning sounding and surface meteorology data as the input. Then, for every feature in the validation set, we disrupt its association with the target by shuffling its values across all instances, creating a permutation of the dataset. This is executed while maintaining the original order of all other features. Furthermore, we recalculate the MAE with the shuffled data. The difference between this new MAE and the baseline MAE represents the feature's importance. To ensure a comprehensive assessment, the permutation and subsequent MAE calculation are repeated 20 times with different random shuffles. The final importance score for each feature is then determined as the mean increase in MAE across these permutations.

However, for the model classifying cloud triggers, which is a classification task, the accuracy score is used as the performance metric. The accuracy score is a measure of the model's overall correctness and is calculated using the formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (1)$$

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

where True Positives (TP) indicates the number of instances correctly predicted as positive; true Negatives (TN) indicates the number of instances correctly predicted as negative; False Positives (FP) indicates the number of instances incorrectly predicted as positive, and False Negatives (FN) indicates the number of instances incorrectly predicted as negative. After determining the performance metric, other procedures for determining feature importance remain the same between regression tasks and the classification task. In the model, we filter individual input parameters from the consideration of importance score.

By using this methodology, Figure 3 illustrates these importance scores from different features, underscoring the most influential factors for predicting the presence, position, and vertical fraction of BLCs. These factors are ranked from most important factors to least important factors. BLC trigger is a special factor since it is the output of the classification model. Notably, the importance scores are not computed as a simple sum but are determined by collectively shuffling groups of features and observing the impact on model performance. Table 1 complements Figure 3 by providing the model's structure and the precise importance values assigned to each feature across the three cloud prediction tasks. Among these factors, LCL is derived from the surface meteorology (Romps, 2017), BLH_{parcel} is derived from the morning temperature profiles and surface air temperature based on the Parcel method (Holzworth, 1964; Su et al. 2020). Specifically, BLH_{parcel} is defined as the height where the morning potential temperature profile first exceeds the current surface potential temperature by more than

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

1.5 K. BLH_{SH} is derived from the morning temperature profiles and surface sensible heat (Stull, 1988; Su et al. 2023).

In particular, surface relative humidity (RH), surface air temperature (T), and morning relative humidity profiles are highly influential in BLC simulations. This is consistent with previous observational and modeling studies (Zhang and Klein, 2013). Surface RH is a critical factor affecting the trigger, CBH, and cloud fraction predictions. As the boundary condition for the DNN model, morning atmospheric profiles of different meteorological parameters (i.e., RH, temperature, and wind components) exert a notable impact on cloud trigger detection and the determination of cloud fractions. Surface air temperature is shown to have a substantial effect on cloud fraction, highlighting the sensitivity of cloud simulations to near-surface thermal conditions. This approach recognizes the interconnectedness of certain features and their collective contribution to the model's output.

4 Modeling Boundary Layer Clouds with Deep Learning

4.1 Boundary Layer Clouds Trigger

The occurrence of BLC is a multifaceted process influenced by a variety of atmospheric parameters and surface processes. The BLC trigger, a critical component in the formation of BLCs, is a dynamic phenomenon that our deep learning model seeks to identify and simulate from the surface meteorology.

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.
Geoscientific Model Development Discussions

By using the morning SONDE and measurements of surface meteorology and fluxes, Figure 4 showcases the model's proficiency in classifying the occurrences (class 1) and non-occurrences (class 0) of BLC during both a trained period and an untrained, future period. The confusion matrices (Luque et al. 2019) for the trained period (1998-2016) and for the untrained period (2017-2020) display the model's predictive performance. The matrices reveal the counts and percentages of TP, FP, TN, and FN. Figure 4a represents the trained period, we use 30% dataset for the validation and see a high percentage of TN at 71.2% and TP at 21.1%, indicating that the model is accurate during the period it was trained on. For the untrained period (2017-2020), the model still performs well, with 71.8% TN and 17.4% TP (Figure 4b). However, the rates of FN and FP are slightly higher at 5.6% and 5.2% respectively, which could indicate that the model is slightly less accurate when applied to data beyond its training scope.

TP, FP, TN, and FN, further offer insights into the model's precision, recall, and overall accuracy in detecting the presence of BLC. Precision, which gauges the accuracy of positive predictions, recall, which assesses the detection of actual positives, and the F1 score, which balances the two, are consistently above 75% across both periods. This high performance across key metrics demonstrates the model's robustness and reliability in identifying the onset of BLCs.

Table 2 complements the Figure 4 and provides a detailed quantitative number of the model's classification performance. It presents the number of instances and their corresponding percentages of different matrices (i.e., TN, FP, FN, and TP). The high

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

percentages of correct predictions (TP and TN) underscore the model's effectiveness, while the lower FP and FN rates reflect its reliability. The table highlights the model's robustness, with overall accuracy rates of 92.3% for the trained period and a slightly reduced but still substantial 89.2% for the untrained period. These metrics demonstrate the model's predictive capabilities and reliability for both trained and untrained periods.

Figure 5 further compares the diurnal frequency of BLC occurrence between observations (OBS) and DNN predictions for different seasons. The BLC's strong diurnal pattern is well-captured by the model, when BLC development peaks between 12-16 local times, aligning closely with observed frequencies. Among different seasons, the model is notably effective in simulating the pronounced diurnal cycle of summer clouds, which are typically influenced by local convection. Conversely, the winter season exhibits a weaker diurnal pattern, likely linking to the diminished surface fluxes. The DNN tends to overestimate BLC presence in the early morning, especially for the winter season. The overall alignment between observations and the DNN model represents the model's capability of capturing the diurnal patterns of BLC formation and development. Determining the triggers of BLC lays the foundation for the integrated simulations of BLC features.

4.2 Simulating Cloud Positions and Cloud Fractions

A key aspect of cloud modeling involves the accurate simulation of cloud positions and fractions, which are indicative of a cloud's vertical extent and fractional coverage.

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

Our deep learning model demonstrates capabilities in predicting these key attributes of BLC.

Figure 6 offer the comparisons between observed values and predictions by the DNN for CBH, CTH, and cloud fraction. These comparisons are presented for both the training period (a, c, e) and an independent period (b, d, f), revealing the model's ability to generalize beyond its initial training data. DNN model demonstrates remarkable performance in simulating cloud base, boasting a correlation coefficient surpassing 0.9 and an MAE under 0.15 km. Conversely, the model encounters challenges with CTH prediction, evidenced by a lower correlation of about 0.5 and a significantly higher MAE between 0.8 and 0.9 km, aligning with case studies in Figure 2.

The discrepancy in accurately simulating CBH and CTH may stem from two main factors. Firstly, CBH determinations are generally more precise due to the effectiveness of laser-based methods (Pal et al., 1992), while CTH estimations often suffer from reduced accuracy, partly attributed to signal attenuation issues (Clothiaux et al., 2000). Secondly, our DNN simulations are developed from the perspective of cloud-land coupling, primarily utilizing surface meteorology. This can introduce inherent limitations, as the tops of many clouds may be decoupled from surface influences despite a coupled base, leading to potential gaps in the model's parameterization to accurately define the cloud top.

The comparison of cloud fraction between observations and DNN are presented to consider the model's capability to simulate the vertical distribution of cloud coverage

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.
Geoscientific Model Development
Discussions

(Figure 6e-f). The scatterplots comparing observed and modeled cloud fractions at individual levels in cloudy scenarios show a satisfactory correlation, with an R-value exceeding 0.77 and an MAE around 0.15. Nevertheless, the DNN model tends to underestimate the peak cloud fraction, displaying a range up to ~0.8 compared to the full 0-1 range observed. This underestimation is intrinsically linked to the model's simulation of cloud position, as both cloud fraction and position models operate in tandem. For stratiform clouds, observational data typically exhibit a relatively uniform vertical extent with cloud fractions close to unity at the central height, whereas the DNN model tends to generate a broader, more attenuated profile with a reduced maximum cloud fraction at the center. This points to a need for refining the model's ability to replicate the pronounced peak cloud fractions characteristic of stratiform cloud profiles.

The diurnal patterns of cloud base and top heights, captured through daily profiles, showcase the model's adeptness at simulating the temporal changes in cloud positions for all BLCs, the cumulus regime, and the stratiform regime (as shown in Figure 7). These profiles, derived from both observational data and DNN outputs, include shaded regions representing the variability (one standard deviation) around the average heights. Cumulus clouds exhibit a marked diurnal cycle, whereas stratiform clouds typically maintain a relatively consistent position and smaller variations throughout the day. A close alignment is observed between the mean and standard deviation of the cloud base between observed and simulated data for different regimes. In contrast, while the mean cloud top heights follow a similar diurnal trend in both cases, the variability presented

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

by the observed data exhibits more pronounced variabilities compared to the relatively small variabilities in DNN simulations.

Figures 6 and 7 collectively demonstrate the model's ability to simulate cloud positions and fractions within BLC. It reliably captures cloud base heights yet encounters challenges with accurately representing cloud top heights and peak cloud fractions on an individual basis. These constraints are somewhat expected, given that even very fine-scale model struggle to entirely capture the vertical extent of clouds, as evidenced in Large-Eddy Simulations or Convection-Permitting Models (Zhang et al. 2017; Gustafson et al. 2020; Bogenschutz et al. 2023).

4.3 Integrating Deep Learning Models into Reanalysis Datasets

The DNN model can use the conventional meteorological data (i.e. morning SONDE and surface meteorology data) to diagnose the BLC. Meanwhile, it also can be used in reanalysis data (i.e., ERA-5 and MERRA-2) to serve as the convection parameterization to simulate BLC with the input of morning profiles and meteorology data from the reanalysis. Thus, we can assess the integration of Deep Learning Models with reanalysis datasets to refine the simulation of BLCs.

Following this thoughts, Figure 8 contrasts diurnal cloud fraction patterns from observational data and deep learning model predictions across all conditions of seasons and years. Figure 8a-b present the observed cloud fractions and those simulated by our deep learning neural network (DNN), respectively. Panels c and e display cloud fractions directly available from ERA and MERRA reanalysis datasets, while panels d

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

and f illustrate the enhanced simulation results after the application of DNN to ERA (ERA_{DNN}) and MERRA ($MERRA_{DNN}$) data. To eliminate sampling biases, we averaged only those samples for which both observational and reanalysis datasets are concurrently available. Observing fluctuations in surface temperature and humidity data in ERA-5 for this region, we smoothed ERA-5 surface air temperature and humidity data with a $\pm$ 1-hour window to mitigate potential variability from assimilation before using them as input for the DNN.

The DNN simulations with observed meteorological data align closely in cloud fraction profiles within the 0-2 km range, reflecting the model's ability to capture land-coupled clouds. As this model are designed for diagnosing land-coupled clouds, the model does not simulate decoupled clouds, which often have bases occurring above the 2-km (Su et al. 2022). Original reanalysis data show significant underestimations of cloud fractions for low clouds, particularly evident in MERRA-2. The implementation of DNN enhances cloud fraction representation compared to original reanalysis data, demonstrating the DNN model's strength in simulating BLC. Given that the DNN model specializes in simulating BLC, when utilizing reanalysis data, the portion of cloud profiles that are decoupled are preserved as they are in the original datasets—that is, for the cloud layers above the BLC-tops or for clouds that rooted above the PBL.

Furthermore, Figure 9 provides a detailed examination of stratiform clouds, utilizing the same comparative approach as in Figure 8. The observed stratiform clouds display a layered structure with expansive coverage and maximum cloud fractions typically exceeding 0.6. The DNN model reproduce these observed characteristics fairly well, albeit with minor overestimations in cloud vertical extent. Conversely, the original ERA-5 and MERRA-2 data exhibit limitations, particularly in underestimating cloud fraction. The integration of the DNN model with reanalysis data enhances the representation of stratiform cloud fractions, as depicted in the heatmaps of Figure 9, showcasing improved agreement with observational data and underscoring the enhancement potential for cloud fraction simulations in reanalysis datasets.

Additionally, Figure 10 extends the comparative study to cumulus clouds. Cumulus clouds pose significant challenges for modeling and parameterization partly due to their typically small spatial extent compared to the model grid, often spanning from a few hundred meters to several kilometers (Zhang et al. 2017; Tao et al., 2021; Bogenschutz et al. 2023; Gustafson et al. 2020). In line with expectations, ERA-5 and MERRA-2 exhibit significant biases in representing cumulus clouds when compared to observational data, possibly related to the large grid of the reanalysis that might not fully capture the fine-scale characteristics of cumulus formations.

In contrast, the DNN model achieves commendable success in capturing the diurnal variability of cumulus clouds, including cloud base, vertical extension, and cloud fraction, by leveraging local convective signals derived from surface meteorology data. When the DNN methodology is applied to ERA-5, it significantly improves the representation of cumulus clouds. However, the original MERRA-2 data, which tend to overlook the majority of cumulus clouds, continue to significantly underrepresent them even after the application of DNN, suggesting that additional biases in meteorological factors may contribute to this discrepancy.

© Author(s) 2024. CC BY 4.0 License.

The integration of deep learning with ERA and MERRA reanalysis datasets demonstrates the notable refinement in the simulation of BLC. By integrating our deep learning models with reanalysis data, we achieve a more accurate representation of cloud fractions for both stratiform and cumulus clouds.

4.4 Applying Deep Learning for Bias Attribution in Cloud Simulation

We further examine the remaining disparities in cloud fraction simulations within reanalysis datasets, despite the integration of deep learning models (as shown in Figures 8-10), indicating persisting meteorological biases. Deep learning is utilized to quantify and attribute these biases for BLC simulations.

Figure 11 offers a comparative analysis of vertical cloud fraction profiles for both stratiform and cumulus clouds. It presents cloud fractions observed with those from reanalysis data (RD), including ERA-5, MERRA-2, and their corresponding deep learning-informed simulations. While the application of deep learning to reanalysis data (RD_{DNN}) yields improvements, remaining biases are evident, particularly in MERRA-2. Acknowledging the significant influence of surface RH on BLC simulations (as indicated by Figure 3e, input observed surface relative humidity (RH) instead of reanalysis RH into the DNN models (the output is RD_{DNN-RH}). This modification leads to a more accurate simulation for MERRA-2, closing the gap with observational data, especially for stratiform clouds. For ERA-5, RD_{DNN-RH} and RD_{DNN} show negligible differences for cumulus clouds, but for stratiform clouds, RD_{DNN-RH} exhibits a reduced bias. These refined profiles attest to the benefits of using observed surface moisture

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

data, confirming its important role in achieving a more accurate representation BLC.

We can further dissect the bias in cloud fraction simulations attributed to various meteorological factors and parameterization schemes within ERA and MERRA reanalysis datasets:

$$Bias\ due\ to\ parameterization = |RD - OBS| - |RD_{DNN} - OBS| \qquad (2)$$

$$Bias\ due\ to\ surface\ RH = |RD_{DNN} - OBS| - |RD_{DNN-RH} - OBS| \qquad (3)$$

where RD and OBS are the cloud fraction derived from reanalysis data and observations, respectively. The definition of RD_{DNN} and RD_{DNN-RH} are the same with the above. For getting a representative value, these biases are layer-averaged from 0-4 km over different local times, and then normalized by the observed mean cloud fraction, offering a climatological perspective on the discrepancies between observed and simulated data across seasons and years.

We get the bias attributed to different meteorological factors and parameterization schemes in the ERA-5 and MERRA-2 datasets, respectively (Figure 12). Each bars indicate the normalized bias contributed by factors such as morning meteorological profiles, surface pressure, surface fluxes, various surface meteorology variables, and parameterization schemes. Notably, parameterization stands out as a significant contributor to bias, accounting for 14.45%/19.05% of the discrepancy in stratiform clouds between observations and ERA-5/MERRA-2. For cumulus clouds, the parameterization biases are more pronounced, contributing 22.23% and 30.94% for ERA-5 and MERRA-2, respectively.

In addition to parameterization, RH, RH profiles, and sensible heat are identified as

© Author(s) 2024. CC BY 4.0 License.

major factors contributing to the differences between observations and reanalysis data. For instance, aligning MERRA-2's RH with observed surface RH could potentially reduce bias by 23.13% for stratiform and 10.26% for cumulus clouds. Meanwhile, surface RH and morning RH profiles in ERA-5 lead to 11.25% and 3.96% of biases for the stratiform clouds. The bias between ERA-5 and observed cumulus clouds is largely driven by parameterization, which suggests that employing the DNN model with ERA-5 can lead to a more accurate simulation of cumulus clouds.

The detailed bias attribution analysis facilitated by the deep learning model elucidates the individual impact of meteorological factors on the discrepancies in cloud fraction between observations and reanalysis data. It underscores the necessity for more accurate humidity data within reanalysis datasets to refine BLC simulations. Furthermore, this deep learning approach illuminates pathways for improved parameterization of boundary layer convection.

5 Summary

This study has developed a deep learning model to estimate the evolution of BLCs over the SGP. The model utilizes over two decades of meteorological data to simulate BLC formation and characteristics, including timing of convection initiation (BLC onset), their positions, and vertical structures. As this model is built based on the perspective of cloud-land coupling, the DNN approach demonstrates the capability to diagnose land-coupled convective systems from surface meteorology. The DNN model is built on the cloud-land interactions and serves as the testimony for the coupling between BLCs and land surface. The proficiency and reliability of the DNN model is evident in its robustness during both the training period and subsequent independent periods.

The implementation of this model within reanalysis datasets like ERA-5 and MERRA-2 has resulted in enhanced representation for stratiform clouds and cumulus and an accurate vertical structure of clouds in term of climatology, providing a promising tool for improving weather forecasting and climate modeling. The deep learning model notably address the limitation in cumulus simulations in the reanalysis data, Meanwhile, this approach is much cost-effective compared to traditional parameterizations and schemes at various scales, as it can simulate two decades of BLC with vertical information over the SGP in ~30-second using a single GPU node.

In addition to the BLC simulations, the deep learning model developed in this study also is used to attribute discrepancies between observational data and reanalysis datasets to different meteorological factors. Besides parameterization, surface RH, morning RH profiles, and surface sensible heat are the three major factors lead to the mismatches in BLC representation in ERA-5 and MERRA-2. These findings underscore the importance of incorporating more accurate humidity information in reanalysis datasets, which is crucial for refining BLC simulations. This analysis also sheds light on the necessity to update reanalysis datasets with improved parameterization of boundary layer convection.

By leveraging deep learning, the model addressed the simulation of cloud vertical structure, among one of the key challenges in the field. They highlight the value of deep

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

learning in advancing our understanding of BLC dynamics and improving the representation of low clouds in atmospheric models. This work not only narrows the observational-modeling divide but also paves the way for future developments in cloud parameterization. Moving forward, future work is warranted to test and extend this parameterization to different synoptic regions. The goal is to develop a versatile model capable of simulating BLC on a global scale, which can be integrated into multiple scale models or reanalysis data.

Code and data availability. The code package of DNN models and for the The simulation outputs of BLCs from observed meteorological data and ERA-5 and MERRA-2 can be found in https://doi.org/10.5281/zenodo.10719342 (Su, 2024). ARM radiosonde data, surface fluxes, and cloud masks are available at https://adc.arm.gov/discovery/#/results/instrument_class_code::armbe (ARM user facility, 1994). ARSCL (Active Remote Sesning of Cloud) can be found in https://adc.arm.gov/discovery/#/results/instrument_class_code::arscl (ARM user facility, 1996). MERRA-2 reanalysis data can be downloaded obtained from https://disc.gsfc.nasa.gov/datasets/M2T1NXRAD_5.12.4/summary?keywords%E2%80%89=%E2%80%89MERRA-2%20tavg1_2d_rad_Nx (GMAO, 2015). ERA-5 reanalysis data are obtained from https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=form (Hersbach et al. 2023).

Author contributions. TS designed this study and carried out the analysis and model training. TS and YZ interpreted the data and wrote the manuscript. YZ supervised the project.

© Author(s) 2024. CC BY 4.0 License.

Competing interests. The contact author has declared that neither they nor their co-authors have any competing interests.

Acknowledgements. Work at LLNL is performed under the auspices of the U.S. DOE by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231. We acknowledge the the U.S. Department of Energy's ARM program for offering the comprehensive filed observations.

Financial support. This work has been supported bythe DOE Atmospheric System Research (ASR) Science Focus Area (SFA) THREAD project.

References

Altmann, A., Toloşi, L., Sander, O. and Lengauer, T. (2010). Permutation importance: a corrected feature importance measure. Bioinformatics, 26(10), pp.1340-1347.

Atmospheric Radiation Measurement (ARM) user facility. 1994. ARM Best Estimate Data Products (ARMBEATM). Southern Great Plains (SGP) Central Facility, Lamont, OK (C1). Compiled by C. Xiao and X. Shaocheng. [Dataset] ARM Data Center. Dataset accessed 2023-12-25 at http://dx.doi.org/10.5439/1333748.

Atmospheric Radiation Measurement (ARM) user facility. 1996. Active Remote Sensing of CLouds (ARSCL1CLOTH). 2024-02-05 to 2024-02-13, Southern Great Plains (SGP) Central Facility, Lamont, OK (C1). Compiled by S. Giangrande, D. Wang, E. Clothiaux and P. Kollias. [Dataset] ARM Data Center. Dataset accessed 2023-12-25 at http://dx.doi.org/10.5439/1996113.

Battaglia, P.W., Hamrick, J.B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V.,

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R. and Gulcehre, C. (2018). Relational inductive biases, deep learning, and graph networks. arXiv preprint arXiv:1806.01261.

Berg, L. K., and Kassianov, E.I. (2008): Temporal variability of fair-weather cumulus statistics at the ACRF SGP site. J. Climate, 21(13), 3344–3358.

Betts, A.K. (2009). Land-surface-atmosphere coupling in observations and models. Journal of Advances in Modeling Earth Systems, 1(3).

Bogenschutz, P. A., Eldred, C., & Caldwell, P. M. (2023). Horizontal resolution sensitivity of the Simple Convection-Permitting E3SM Atmosphere Model in a doubly-periodic configuration. Journal of Advances in Modeling Earth Systems, 15, e2022MS003466. https://doi.org/10.1029/2022MS003466

Bretherton, C. S., Blossey, P. N., and Uchida, J. (2007): Cloud droplet sedimentation, entrainment efficiency, and subtropical stratocumulus albedo. Geophys. Res. Lett., 34(3), L03813.

Cadeddu, M. P., Turner, D. D., and Liljegren, J. C. (2009). A neural network for real-time retrievals of PWV and LWP from Arctic millimeter-wave ground-based observations, IEEE T. Geosci. Remote, 47, 1887 – 1900.

Caldwell, P.M., Terai, C.R., Hillman, B., Keen, N.D., Bogenschutz, P., Lin, W., Beydoun, H., Taylor, M., Bertagna, L., Bradley, A.M. and Clevenger, T.C., 2021. Convection-permitting simulations with the E3SM global atmosphere model. Journal of Advances in Modeling Earth Systems, 13(11), p.e2021MS002544.

Clothiaux, E. E., Ackerman, T. P., Mace, G. G., Moran, K. P., Marchand, R. T., Miller, M. A., and Martner, B. E. (2000): Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the ARM CART sites. J. Appl. Meteorol., 39(5), 645 – 665.

Clothiaux, E.E., Miller, M.A., Perez, R.C., Turner, D.D., Moran, K.P., Martner, B.E., Ackerman, T.P., Mace, G.G., Marchand, R.T., Widener, K.B. and Rodriguez, D.J. (2001). The ARM millimeter wave cloud radars (MMCRs) and the active remote sensing of clouds (ARSCL) value added product (VAP) (No. DOE/SC-ARM/VAP-002.1). DOE Office of Science Atmospheric Radiation Measurement (ARM) Program (United States).

Cook, D. R. (2018). Energy Balance Bowen Ratio (EBBR) instrument handbook, Technical Report Rep. DOE/SC-ARM/TR-037, U.S. Department of Energy.

Date, Y. and Kikuchi, J. (2018). Application of a deep neural network to metabolomics studies and its performance in determining important variables. Analytical chemistry, 90(3), pp.1805-1810.

Fast, J. D., Berg, L. K., Alexander, L., Bell, D., D'Ambro, E., Hubbe, J., Kuang, C., Liu, J., Long, C., Matthews, A., and Mei, F.: Overview of the HI-SCALE field campaign: A new perspective on shallow convective clouds, B. Am. Meteorol. Soc., 100, 821 – 840, (2019).

Gagne II, D. J., Haupt, S. E., Nychka, D. W., and Thompson, G.: Interpretable deep learning for spatial analysis of severe hailstorms, Mon. Weather Rev., 147, 2827 – 2845, (2019).

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R. and Wargan, K., 2017. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). Journal of climate, 30(14), pp.5419-5454.

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G. and Yacalis, G., 2018. Could machine learning break the convection parameterization deadlock?. Geophysical Research Letters, 45(11), pp.5742-5751.

Global Modeling and Assimilation Office (GMAO) (2015), MERRA-2 tavg1_2d_rad_Nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Radiation Diagnostics V5.12.4, Greenbelt, MD, USA [Dataset]. Goddard Earth Sciences Data and Information Services Center (GES DISC), https://doi.org/10.5067/Q9QMY5PBNV1T.

Golaz, J.C., Larson, V.E. and Cotton, W.R. (2002). A PDF-based model for boundary layer clouds. Part I: Method and model description. Journal of the atmospheric sciences, 59(24), pp.3540-3551.

Guo, J., Su, T., Chen, D., Wang, J., Li, Z., Lv, Y., ... & Zhai, P. (2019). Declining summertime local‐scale precipitation frequency over China and the United States, 1981–2012: The disparate roles of aerosols. Geophysical Research Letters, 46(22), 13281-13289.

Guo, J., Zhang, J., Shao, J., Chen, T., Bai, K., Sun, Y., Li, N., Wu, J., Li, R., Li, J. and Guo, Q., 2024. A merged continental planetary boundary layer height dataset based on high-resolution radiosonde measurements, ERA5 reanalysis, and GLDAS. Earth System Science Data, 16(1), pp.1-14.

Gustafson, W.I., Vogelmann, A.M., Li, Z., Cheng, X., Dumas, K.K., Endo, S., Johnson, K.L., Krishna, B., Fairless, T. and Xiao, H., 2020. The large-eddy simulation (LES) atmospheric radiation measurement (ARM) symbiotic simulation and observation (LASSO) activity for continental shallow convection. Bulletin of the American Meteorological Society, 101(4), pp.E462-E479.

Haynes, J.M., Noh, Y.J., Miller, S.D., Haynes, K.D., Ebert-Uphoff, I. and Heidinger, A., 2022. Low cloud detection in multilayer scenes using satellite imagery with machine learning methods. Journal of Atmospheric and Oceanic Technology, 39(3), pp.319-334.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J-N. (2023): ERA5 hourly data on pressure levels from 1940 to present. [Dataset] Copernicus Climate Change Service (C3S) Climate Data Store (CDS), DOI: 10.24381/cds.bd0915c6.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D. and Simmons, A., 2020. The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), pp.1999-2049.

Holdridge, D., Ritsche, M., Prell, J., and Coulter, R. (2011): Balloon-borne sounding system (SONDE) handbook, https://www.arm.gov/capabilities/instruments/sonde.

Holzworth, G. C.: Estimates of mean maximum mixing depths in the contiguous United States, Mon. Weather Rev., 92, 235–242, https://doi.org/10.1175/1520-

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

Geoscientific Model Development Discussions

0493(1964)092<0235:eommmd>2.3.co;2, 1964.

Kollias, P., Bharadwaj, N., Clothiaux, E.E., Lamer, K., Oue, M., Hardin, J., Isom, B., Lindenmaier, I., Matthews, A., Luke, E.P. and Giangrande, S.E. (2020). The ARM radar network: At the leading edge of cloud and precipitation observations. Bulletin of the American Meteorological Society, 101(5), pp.E588-E607.

Kuma, P., McDonald, A.J., Morgenstern, O., Alexander, S.P., Cassano, J.J., Garrett, S., Halla, J., Hartery, S., Harvey, M.J., Parsons, S. and Plank, G., 2020. Evaluation of Southern Ocean cloud in the HadGEM3 general circulation model and MERRA-2 reanalysis using ship-based observations. Atmospheric Chemistry and Physics, 20(11), pp.6607-6630.

Lareau, N.P., Zhang, Y. and Klein, S.A. (2018). Observed boundary layer controls on shallow cumulus at the ARM Southern Great Plains site. Journal of the Atmospheric Sciences, 75(7), pp.2235-2255.

Lee, J.M., Zhang, Y. and Klein, S.A. (2019). The effect of land surface heterogeneity and background wind on shallow cumulus clouds and the transition to deeper convection. Journal of the Atmospheric Sciences, 76(2), pp.401-419.

Lu, C., Liu, Y. and Niu, S., 2011. Examination of turbulent entrainment‐mixing mechanisms using a combined approach. Journal of Geophysical Research: Atmospheres, 116(D20).

Lu, C., Niu, S., Liu, Y. and Vogelmann, A.M., 2013. Empirical relationship between entrainment rate and microphysics in cumulus clouds. Geophysical Research Letters, 40(10), pp.2333-2338.

Luque, A., Carrasco, A., Martín, A. and de Las Heras, A., 2019. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. Pattern Recognition, 91, pp.216-231.

McGovern, A., Elmore, K. L., Gagne, D. J., Haupt, S. E., Karstens, C. D., Lagerquist, R., Smith, T., and Williams, J. K.: Using artificial intelligence to improve real-time decision-making for high-impact weather, B. Am. Meteorol. Soc., 98, 2073–2090, (2017).

Miao, H., Wang, X., Liu, Y. and Wu, G. (2019). An evaluation of cloud vertical structure in three reanalyses against CloudSat/cloud‐aerosol lidar and infrared pathfinder satellite observations. Atmospheric Science Letters, 20(7), p.e906.

Moeng, C.H., Cotton, W.R., Bretherton, C., Chlond, A., Khairoutdinov, M., Krueger, S., Lewellen, W.S., MacVean, M.K., Pasquier, J.R.M., Rand, H.A. and Siebesma, A.P. (1996). Simulation of a stratocumulus-topped planetary boundary layer: Intercomparison among different numerical codes. Bulletin of the American Meteorological Society, 77(2), pp.261-278.

Molero, F., Barragán, R. and Artíñano, B. (2022). Estimation of the atmospheric boundary layer height by means of machine learning techniques using ground-level meteorological data. Atmospheric Research, 279, p.106401.

Mooers, G., Pritchard, M., Beucler, T., Ott, J., Yacalis, G., Baldi, P. and Gentine, P., 2021. Assessing the potential of deep learning for emulating cloud superparameterization in climate models with real‐geography boundary conditions. Journal of Advances in Modeling Earth Systems, 13(5),

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

p.e2020MS002385.

Nogherotto, R., Tompkins, A.M., Giuliani, G., Coppola, E. and Giorgi, F., 2016. Numerical framework and performance of the new multiple-phase cloud microphysics scheme in RegCM4. 5: precipitation, cloud microphysics, and cloud radiative effects. Geoscientific Model Development, 9(7), pp.2533-2547.

O'Gorman, P.A. and Dwyer, J.G., 2018. Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. Journal of Advances in Modeling Earth Systems, 10(10), pp.2548-2563.

Pal, S.R., Steinbrecht, W. and Carswell, A.I., 1992. Automated method for lidar determination of cloud-base height and vertical extent. Applied optics, 31(10), pp.1488-1494.

Poll, S., Shrestha, P. and Simmer, C. (2022). Grid resolution dependency of land surface heterogeneity effects on boundary-layer structure. Quarterly Journal of the Royal Meteorological Society, 148(742), pp.141-158.

Qian, Y., Guo, Z., Larson, V.E., Leung, L.R., Lin, W., Ma, P.L., Wan, H., Wang, H., Xiao, H., Xie, S. and Yang, B. (2023). Region and cloud regime dependence of parametric sensitivity in E3SM atmosphere model. Climate Dynamics, pp.1-17.

Raju, V.G., Lakshmi, K.P., Jain, V.M., Kalidindi, A. and Padma, V. (2020), August. Study the influence of normalization/transformation process on the accuracy of supervised classification. In 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 729-735). IEEE.

Randles, C.A., Da Silva, A.M., Buchard, V., Colarco, P.R., Darmenov, A., Govindaraju, R., Smirnov, A., Holben, B., Ferrare, R., Hair, J. and Shinozuka, Y. (2017). The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation. Journal of climate, 30(17), pp.6823-6850.

Rasp, S., 2020. Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations: general algorithms and Lorenz 96 case study (v1. 0). Geoscientific Model Development, 13(5), pp.2185-2196.

Rieck, M., Hohenegger, C. and van Heerwaarden, C.C. (2014). The influence of land surface heterogeneities on cloud size development. Monthly Weather Review, 142(10), pp.3830-3846.

Romps, D.M. (2017). Exact expression for the lifting condensation level. Journal of the Atmospheric Sciences, 74(12), pp.3891-3900.

Sakaguchi, K., Berg, L.K., Chen, J., Fast, J., Newsom, R., Tai, S.L., Yang, Z., Gustafson Jr, W.I., Gaudet, B.J., Huang, M. and Pekour, M. (2022). Determining spatial scales of soil moisture—Cloud coupling pathways using semi‐idealized simulations. Journal of Geophysical Research: Atmospheres, 127(2), p.e2021JD035282.

Stull, R.B. (1988). An Introduction to Boundary Layer Meteorology. Dordrecht: Springer Netherlands

Su, T. (2024). Codes and Package of Deep Learning Driven Simulations of Boundary Layer Cloud over the US Southern Great Plains [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.10685605

Su, T., Li, Z. and Zheng, Y. (2023). Cloud‐Surface Coupling Alters the Morning Transition From Stable to Unstable Boundary Layer. Geophysical Research Letters,

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

50(5), p.e2022GL102256.

Su, T., Li, Z., and Kahn, R. (2020): A new method to retrieve the diurnal variability of planetary boundary layer height from lidar under different thermodynamic stability conditions. Remote Sens. Environ., 237, 111519.

Su, T., Zheng, Y. and Li, Z. (2022): Methodology to determine the coupling of continental clouds with surface from lidar and meteorological data. Atmos. Chem. Phys.

Tang, Q., Xie, S., Zhang, Y., Phillips, T. J., Santanello, J. A., Cook, D. R., Riihimaki, L. D., and Gaustad, K. L.: Heterogeneity in warm-season land-atmosphere coupling over the US Southern Great Plains, J. Geophys. Res.-Atmos., 123, 7867–7882, (2018).

Tang, S., Xie, S., Zhang, M., Tang, Q., Zhang, Y., Klein, S. A., Cook, D. R., and Sullivan, R. C. (2019): Differences in eddy-correlation and energy-balance surface turbulent heat flux measurements and their impacts on the large-scale forcing fields at the ARM SGP site. J. Geophy. Res. Atmos., 124, 3301–3318, doi.org/10.1029/2018JD029689.

Tao, C., Zhang, Y., Tang, Q., Ma, H.Y., Ghate, V.P., Tang, S., Xie, S. and Santanello, J.A. (2021). Land–Atmosphere coupling at the US Southern Great Plains: A comparison on local convective regimes between ARM observations, reanalysis, and climate model simulations. Journal of Hydrometeorology, 22(2), pp.463-481.

Tao, C., Zhang, Y., Tang, S., Tang, Q., Ma, H.Y., Xie, S. and Zhang, M. (2019). Regional moisture budget and land-atmosphere coupling over the US Southern Great Plains inferred from the ARM long-term observations. Journal of Geophysical Research: Atmospheres, 124(17-18), pp.10091-10108.

Teixeira, J., and Hogan, T. F. (2002): Boundary layer clouds in a global atmospheric model: simple cloud cover parameterizations. J. Climate, 15(11), 1261–1276.

Tian, J., Zhang, Y., Klein, S.A., Öktem, R. and Wang, L. (2022). How does land cover and its heterogeneity length scales affect the formation of summertime shallow cumulus clouds in observations from the US Southern Great Plains?. Geophysical Research Letters, 49(7), p.e2021GL097070.

Tiedtke, M. (1993). Representation of clouds in large-scale models. Monthly Weather Review, 121(11), 3040-3061.

Vassallo, D., Krishnamurthy, R., and Fernando, H. J. S.: Decreasing wind speed extrapolation error via domain-specific feature extraction and selection, Wind Energ. Sci., 5, 959–975, https://doi.org/10.5194/wes-5-959-2020, (2020).

Wang, C., Platnick, S., Meyer, K., Zhang, Z. and Zhou, Y., 2020. A machine-learning-based cloud detection and thermodynamic-phase classification algorithm using passive spectral observations. Atmospheric Measurement Techniques, 13(5), pp.2257-2277.

Wang, Y., Zhang, R. and Saravanan, R., 2014. Asian pollution climatically modulates mid-latitude cyclones following hierarchical modelling and observational analysis. Nature communications, 5(1), p.3098.

Wang, Y., Zheng, X., Dong, X., Xi, B. and Yung, Y.L., 2023. Insights of warm-cloud biases in Community Atmospheric Model 5 and 6 from the single-column modeling

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

framework and Aerosol and Cloud Experiments in the Eastern North Atlantic (ACE-ENA) observations. Atmospheric Chemistry and Physics, 23(15), pp.8591-8605.

Wang, Y., Zheng, X., Dong, X., Xi, B., Wu, P., Logan, T. and Yung, Y.L., 2020. Impacts of long-range transport of aerosols on marine-boundary-layer clouds in the eastern North Atlantic. Atmospheric Chemistry and Physics, 20(23), pp.14741-14755.

Xiao, H., Berg, L.K. and Huang, M. (2018). The impact of surface heterogeneities and land-atmosphere interactions on shallow clouds over ARM SGP site. Journal of Advances in Modeling Earth Systems, 10(6), pp.1220-1244.

Xiao, H., Berg, L.K. and Huang, M. (2018). The impact of surface heterogeneities and land-atmosphere interactions on shallow clouds over ARM SGP site. Journal of Advances in Modeling Earth Systems, 10(6), pp.1220-1244.

Xie, S., McCoy, R. B., Klein, S. A., Cederwall, R. T., Wiscombe, W. J., Jensen, M. P., Johnson, K. L., Clothiaux, E. E., Gaustad, K. L., Long, C. N., and Mather, J. H. (2010): Clouds and more: ARM climate modeling best estimate data: a new data product for climate studies. Bull. Amer. Meteorol. Soc., 91(1), 13 - 20.

Yang, Y., Zheng, X., Gao, Z., Wang, H., Wang, T., Li, Y., Lau, G.N. and Yim, S.H., 2018. Long - term trends of persistent synoptic circulation events in planetary boundary layer and their relationships with haze pollution in winter half year over eastern China. Journal of Geophysical Research: Atmospheres, 123(19), pp.10-991.

Yeo, H., Kim, M.H., Son, S.W., Jeong, J.H., Yoon, J.H., Kim, B.M. and Kim, S.W., 2022. Arctic cloud properties and associated radiative effects in the three newer reanalysis datasets (ERA5, MERRA-2, JRA-55): Discrepancies and possible causes. Atmospheric Research, 270, p.106080.

Zhang, T., Lin, W., Vogelmann, A.M., Zhang, M., Xie, S., Qin, Y. and Golaz, J.C., 2021. Improving convection trigger functions in deep convective parameterization schemes using machine learning. Journal of Advances in Modeling Earth Systems, 13(5), p.e2020MS002365.

Zhang, Y. and Klein, S.A. (2010). Mechanisms affecting the transition from shallow to deep convection over land: Inferences from observations of the diurnal cycle collected at the ARM Southern Great Plains site. Journal of the Atmospheric Sciences, 67(9), pp.2943-2959.

Zhang, Y. and Klein, S.A., 2013. Factors controlling the vertical extent of fair-weather shallow cumulus clouds over land: Investigation of diurnal-cycle observations collected at the ARM Southern Great Plains site. Journal of the Atmospheric Sciences, 70(4), pp.1297-1315.

Zhang, Y., Klein, S.A., Fan, J., Chandra, A.S., Kollias, P., Xie, S. and Tang, S. (2017). Large-eddy simulation of shallow cumulus over land: A composite case based on ARM long-term observations at its Southern Great Plains site. Journal of the Atmospheric Sciences, 74(10), pp.3229-3251.

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.
Geoscientific Model Development Discussions
Open Access EGU

TABLE LIST:

Table 1: Feature Importance and Structure in the Deep Learning Models. This table details the structure and importance of features for each DNN model applied to predict cloud trigger, cloud base height (CBH), and cloud fraction. The structure of each model is defined by the number of neurons in sequential layers, and the importance scores reflect the predictive contribution of each feature, which includes month, local time (LT), surface pressure (PS), relative humidity (RH), zonal (U) and meridional (V) wind components, temperature (T), lifting condensation level (LCL), boundary layer height derived from sensible heat (BLH_{SH}) and parcel methods (BLH_{Parcel}), sensible heat (SH), latent heat (LH), and morning profiles of relative humidity (RH Profile), U wind (U Profile), V wind (V Profile), and potential temperature (θ Profile). The trigger feature is applicable only to the CBH and cloud fraction models.

Feature		**Trigger**	**CBH**	**Cloud fraction**
Structure		[108, 64, 36, 24]	[96, 56, 32, 24]	[56, 32, 24]
Importance	**Month**	0.00527286	0.001389318	0.005121363
	LT	0.02753033	0.012310212	0.012221336
	PS	0.004575528	0.00105115	0.004330988
	RH	0.107300553	0.54430006	0.06879726
	U	0.007313414	0.002259621	0.003518908
	V	0.025564082	0.012099721	0.00784819
	T	0.006910447	0.105152747	0.073418902
	LCL	0.00455838	-0.001687665	-0.000544992
	BLH_{SH}	0.006687529	0.010036585	0.008999221
	BLH_{Parcel}	0.056904017	0.088704497	0.014904008
	SH	0.006364585	0.008868474	0.016607784
	LH	0.002783613	0.010357209	0.024233806
	RH Profile	0.097609351	0.02356447	0.050633957
	U Profile	0.030436833	0.010810712	0.04678593
	V Profile	0.049247653	0.023581766	0.045574382
	θ Profile	0.024626684	0.015502336	0.034399533
	Trigger	N.A.	0.017457138	0.095616528

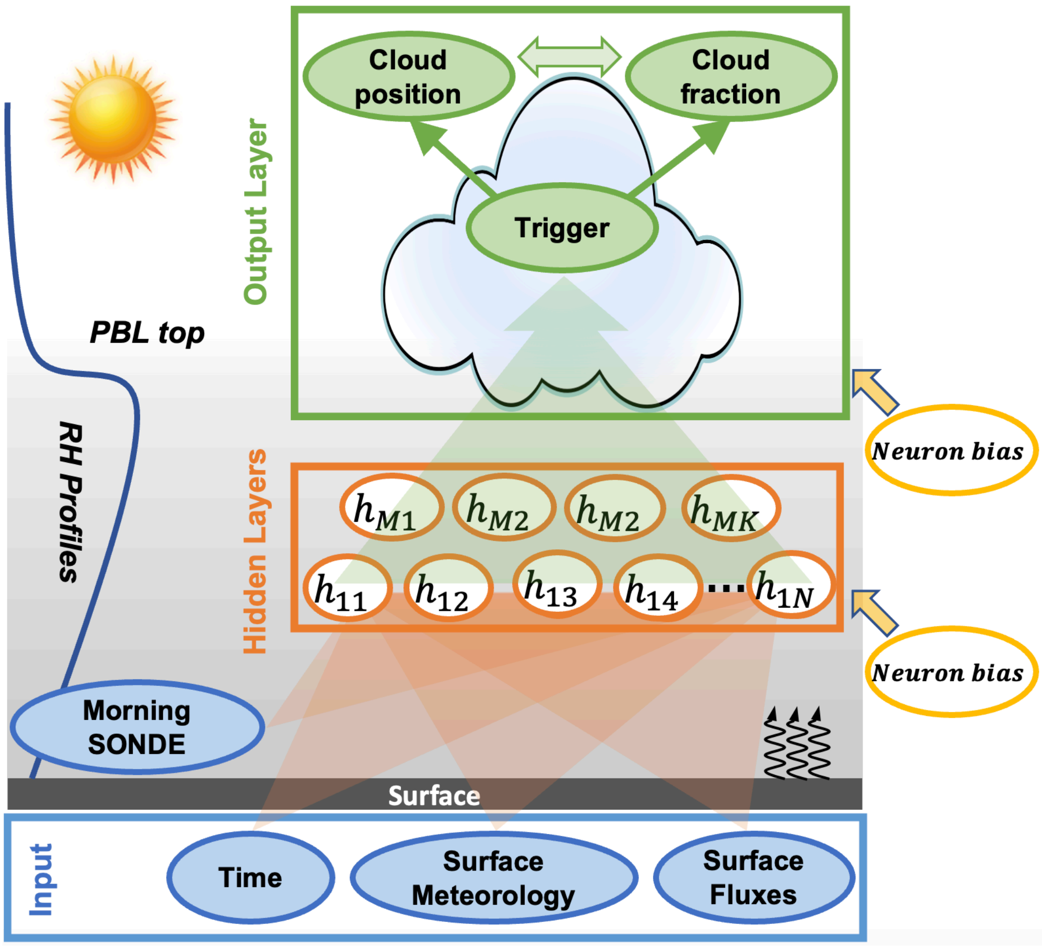
https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

Table 2: Classification Performance of the Deep Learning Model for Boundary Layer Clouds (BLC) Trigger. This table present the performance metrics of the deep learning model during both the trained and untrained periods. It lists the number of samples and corresponding percentages for true negatives (TN), false positives (FP), false negatives (FN), and true positives (TP). The overall accuracy for each period is also provided, indicating the model's overall effectiveness in predicting the presence of boundary layer clouds.

Performance Metrics	**Trained Period (1998-2016)**		**Untrained Period (2017-2020)**	
	Sample #	**Percentage (%)**	**Sample #**	**Percentage (%)**
TN	9773	71.1747142	5416	71.8016704
FP	393	2.8621368	393	5.2101286
FN	670	4.8794698	424	5.6211057
TP	2895	21.0836793	1310	17.3670953
Overall Accuracy	N.A.	92.2583934	N.A.	89.1687657

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

914 **Figures**

915 

916 **Figure 1:** Conceptual diagram of the deep learning framework for simulating boundary
917 layer cloud (BLC) characteristics over the US Southern Great Plains. Inputs for the deep
918 neural networks (DNN) include morning radiosonde (SONDE) profiles, time indicators,
919 and surface conditions such as fluxes (curved black arrows) and meteorological data.
920 The relevance of relative humidity (RH) profiles and the planetary boundary layer (PBL)
921 top is emphasized due to their critical role in boundary layer development. These
922 variables are processed through multiple layers of hidden neurons (h_{11} to h_{MK}), each
923 with neuron bias adjustments to optimize the network's predictive capability. Separate
924 DNN models are constructed for the initiation (trigger) of boundary layer clouds (BLC),
925 their vertical positioning, and cloud fraction across ten atmospheric layers. Together,
926 these models synergize to predict the presence, altitude, and stratification of BLC.

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.
Geoscientific Model Development
Discussions

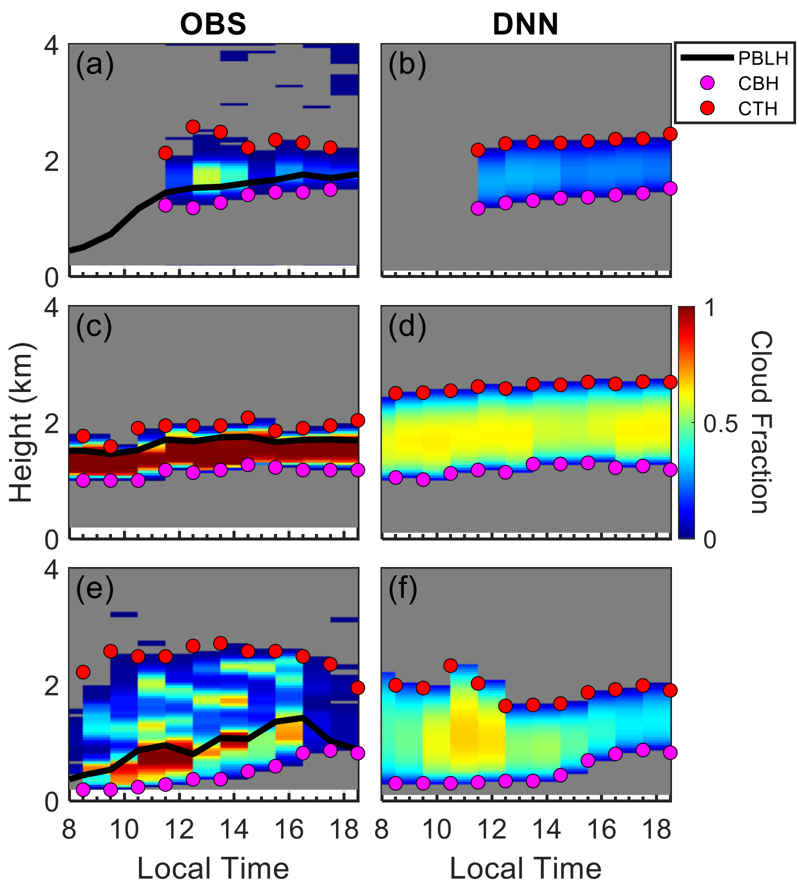


Figure 2: Examples of diurnal cloud fraction profiles for cumulus (a, b), stratiform (c, d), and complex cloud structures (e, f) over the US Southern Great Plains. Observed data (OBS) are shown alongside deep learning neural network (DNN) simulations. Black lines represent the observed PBL height (PBLH), with cloud base (CBH) and cloud top heights (CTH) marked by pink and red dots, respectively. The color gradient indicates the cloud fraction.

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.
Geoscientific Model Development
Open Access
Discussions

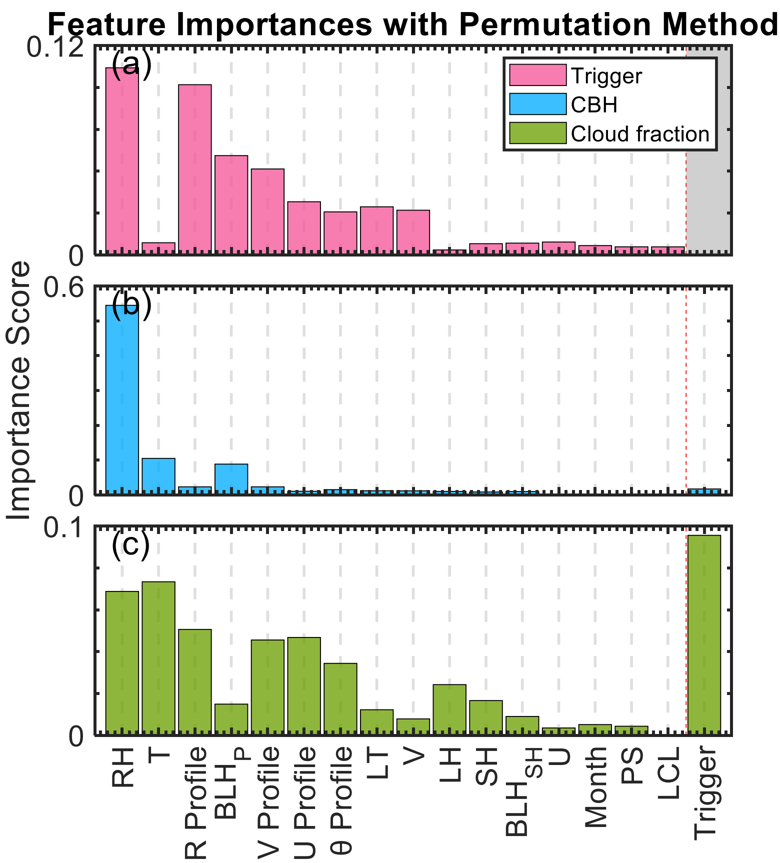


Figure 3. Feature importance scores for predicting cloud trigger (a), cloud base height (CBH) (b), and cloud fraction (c) in deep learning simulations of BLCs. Each panel presents the relative contribution of input features, includes month, local time (LT), surface pressure (PS), relative humidity (RH), zonal (U) and meridional (V) wind components, temperature (T), lifting condensation level (LCL), boundary layer height derived from sensible heat (BLH_{SH}) and parcel methods (BLH_{Parcel}), sensible heat (SH), latent heat (LH), and morning profiles of relative humidity (R Profile), U wind (U Profile), V wind (V Profile), and potential temperature (θ Profile). These factors are ranked based on their overall importance. The importance scores are calculated with permutation method and quantify the relative contribution of each feature to the model's predictive accuracy.

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

Geoscientific Model Development
Discussions
Open Access

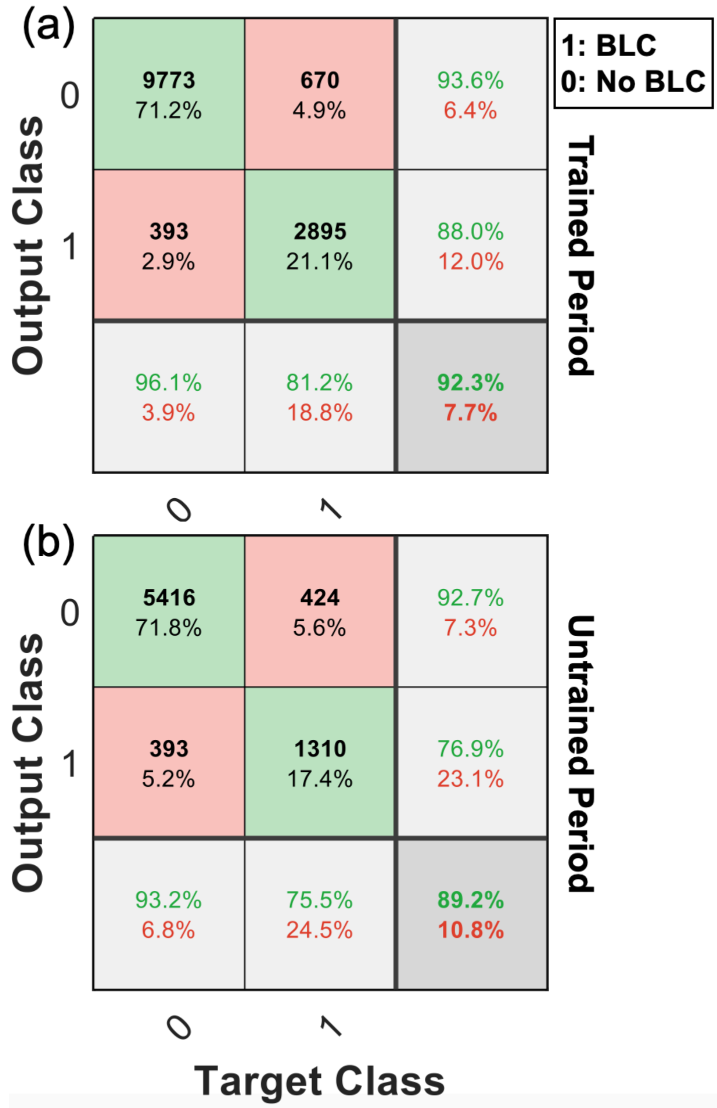


Figure 4: Confusion matrices representing the classification performance of the deep learning model for the presence of boundary layer clouds (BLCs) during the trained period (1998-2016) in panel (a), and the untrained period (2017-2020) in panel (b). For the trained period, we use 30% dataset for the validation. The matrices in the black color display the counts and percentages of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) predictions. The overall precision, recall, and F1 scores for each class are also included, demonstrating the model's ability in identifying BLC occurrence.

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

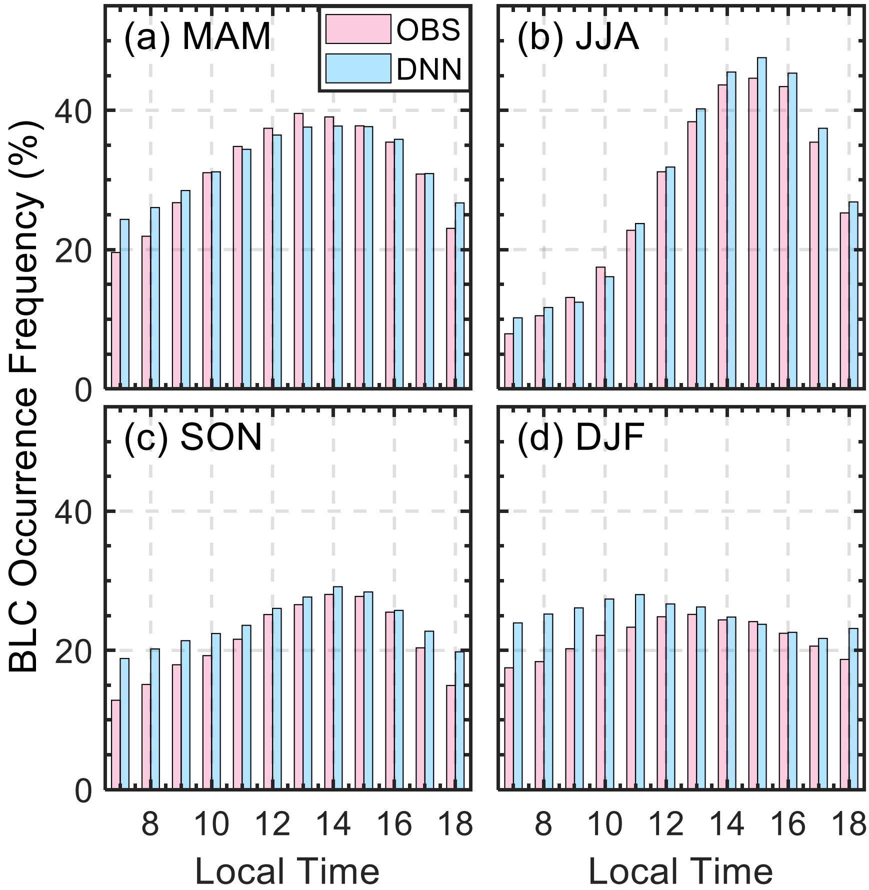


Figure 5. Bar graph comparing the frequency of boundary layer cloud (BLC) occurrence as observed (OBS, red) and as predicted by the deep learning neural network (DNN, blue) across different local times of the day, segmented by seasons: (a) MAM (Spring), (b) JJA (Summer), (c) SON (Fall), and (d) DJF (Winter). The bars present the diurnal pattern of BLC development.

© Author(s) 2024. CC BY 4.0 License.

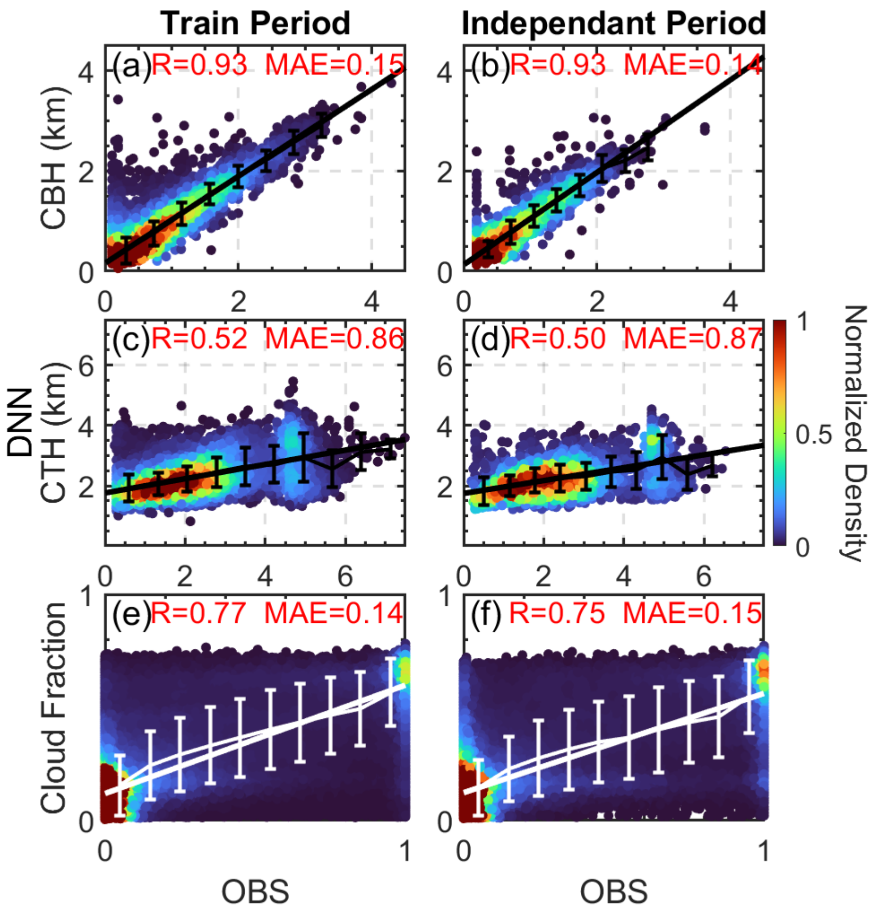


Figure 6. Scatter density plots comparing observed (OBS) and deep learning neural network (DNN) predicted values for cloud base height (CBH), cloud top height (CTH), and cloud fraction during the train period (a, c, e) and an independent period (b, d, f). Note that the BLC is segmented into ten layers, yielding ten separate cloud fraction values per BLC instance for analysis. The correlation coefficient (R) and mean absolute error (MAE) are indicated for each comparison. The color scale represents the normalized density of data points. The solid lines and error bars denoting the linear regression and standard deviations in each bar.

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.

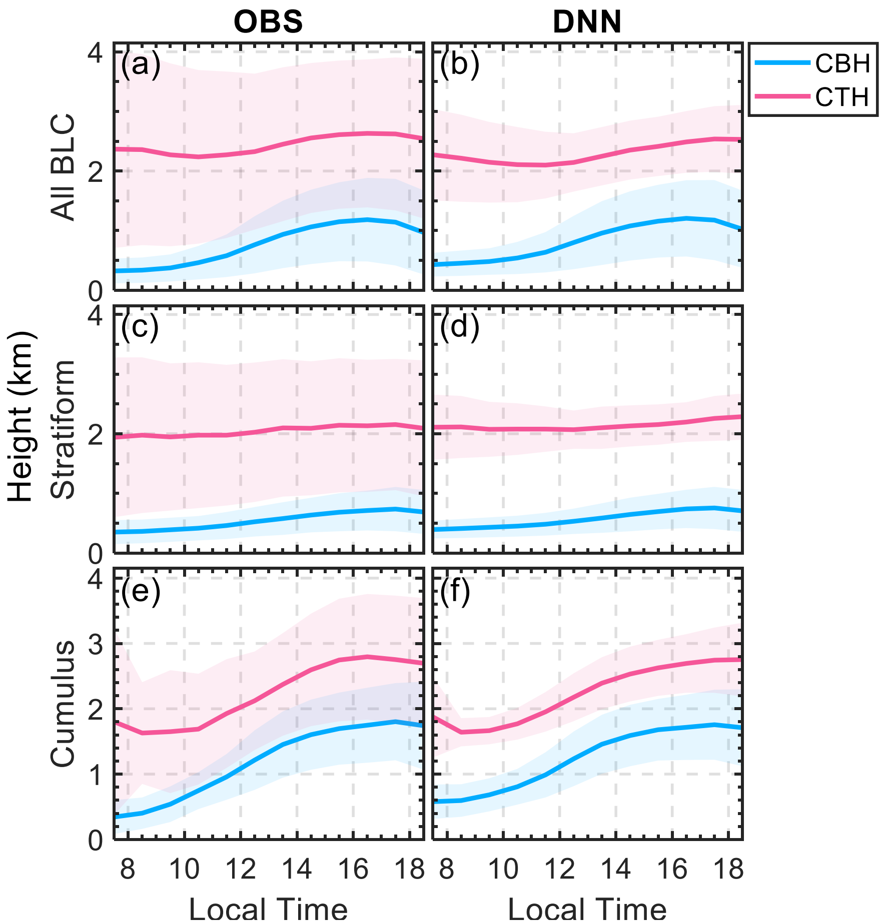
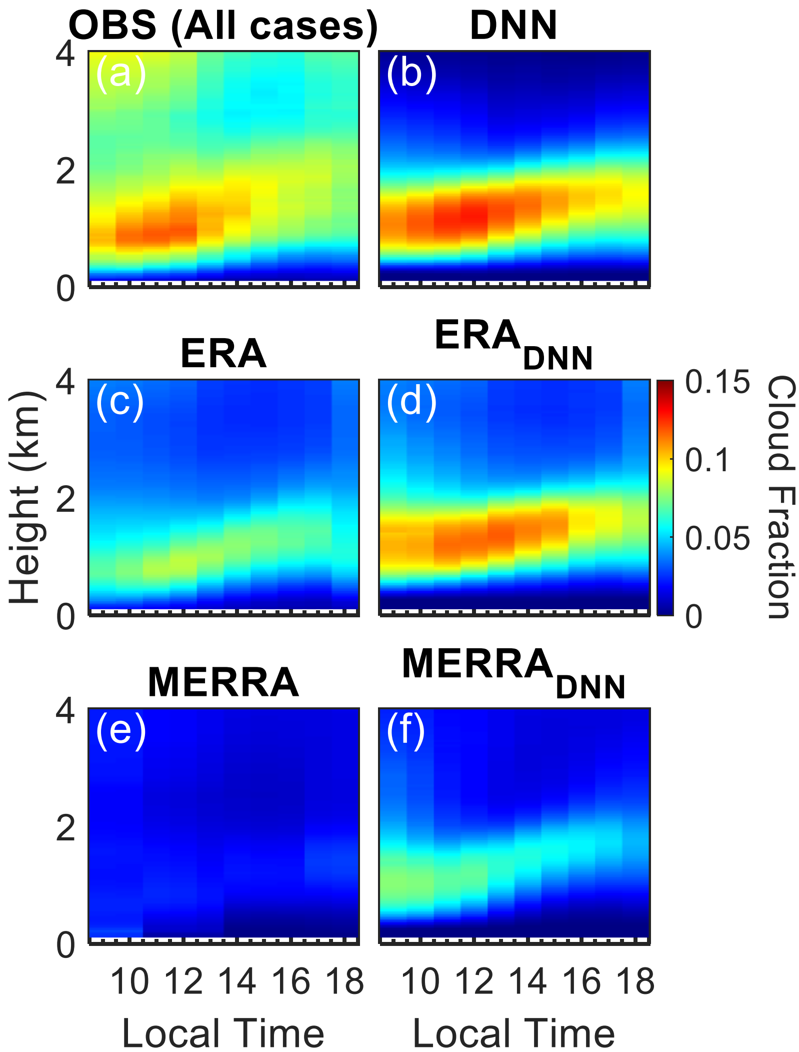


Figure 7. Diurnal profiles of cloud base height (CBH) and cloud top height (CTH) as determined by the observations (OBS) and deep learning simulations for all BLC (a-b), stratiform clouds (c-d), and cumulus (e-f). The shaded areas represent the variability (one standard deviation) around the mean heights.

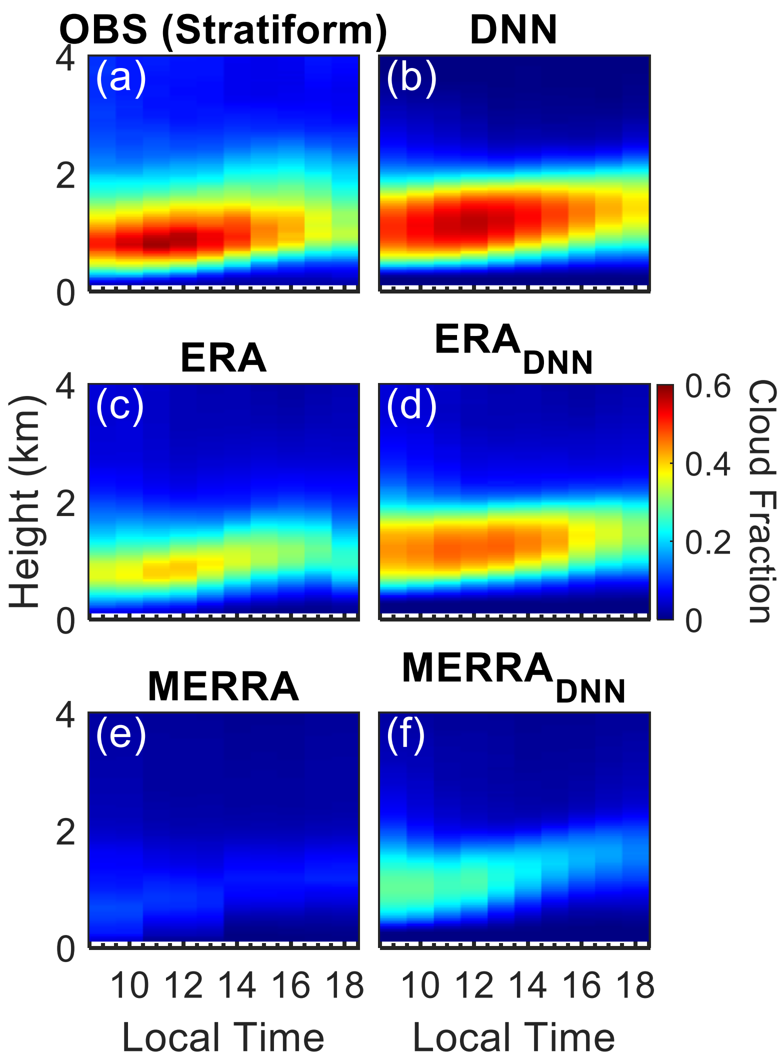
https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.
Geoscientific Model Development
Discussions
Open Access
EGU



980

981 **Figure 8.** Cloud shaded areas demonstrate the diurnal variation in cloud fraction for all
982 cases as observed and simulated. Panel (a) shows the observed cloud fraction (OBS),
983 while panel (b) illustrates the cloud fraction simulated by the deep learning neural
984 networks (DNN). (c, e): cloud fractions from ERA and MERRA reanalysis datasets,
985 respectively. (d, f): the cloud fractions after the application of the DNN model to ERA
986 (ERA$_{DNN}$) and MERRA (MERRA$_{DNN}$) data.

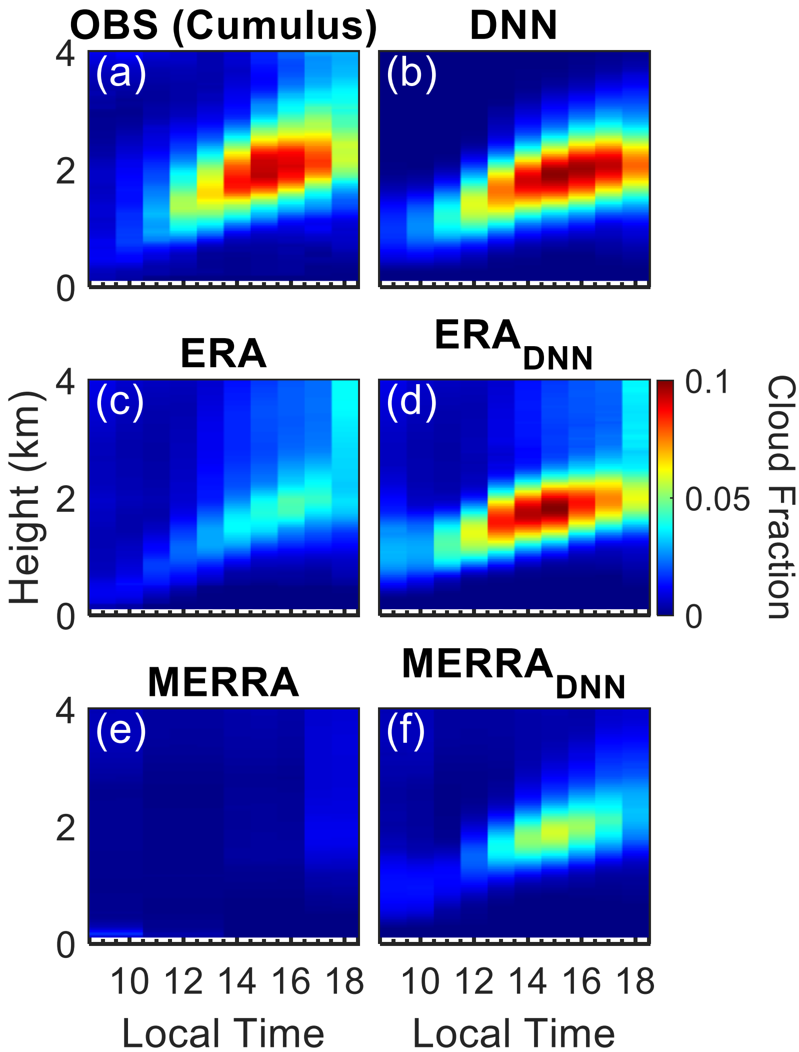
https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.
Geoscientific Model Development Discussions



987

988 **Figure 9.** Same to Figure 8, but for stratiform clouds.

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.
Geoscientific Model Development
Open Access
Discussions



989

990 **Figure 10.** Same to Figure 8, but for cumulus.

991

https://doi.org/10.5194/gmd-2024-25
Preprint. Discussion started: 11 March 2024
© Author(s) 2024. CC BY 4.0 License.
Geoscientific Model Development
Discussions

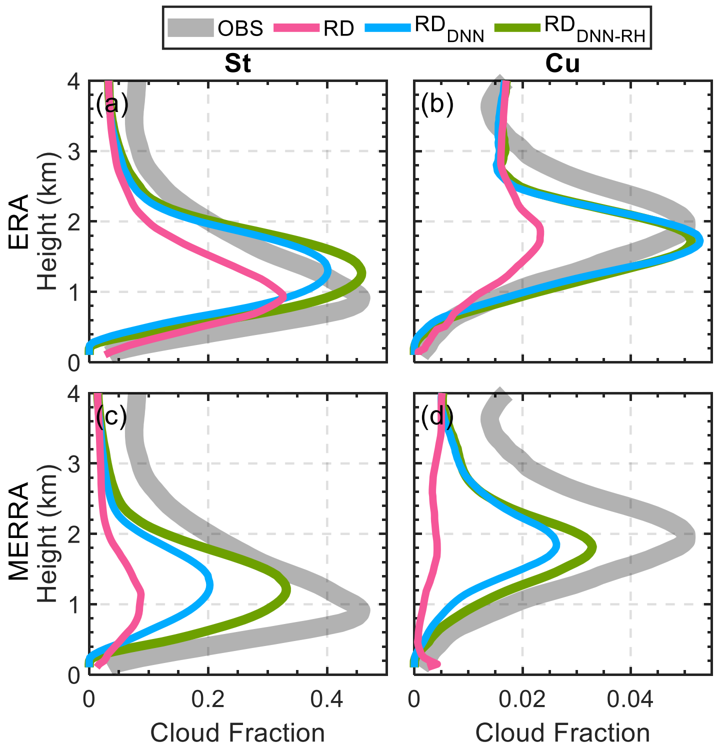


Figure 11: Vertical profiles of cloud fraction for stratiform (St) and cumulus (Cu) scenarios over the US Southern Great Plains. Panels (a) and (b) display ERA reanalysis data comparisons, while panels (c) and (d) show MERRA reanalysis data comparisons. The observed cloud fractions (OBS) are represented by the shaded grey area, illustrating the averaged cloud coverage recorded by field observations. The original reanalysis data (RD) is indicated in pink, indicating the baseline cloud fraction profiles as simulated by the reanalysis. The RD$_{DNN}$ profiles in blue depict the new simulation results after applying the DNN models to the reanalysis data for boundary layer cloud (BLC) simulation. The RD$_{DNN-RH}$ profiles in green show the simulation results when the surface relative humidity (RH) from the reanalysis data is replaced with observed values, indicating the impact of accurate surface moisture representation on cloud fraction simulations.

© Author(s) 2024. CC BY 4.0 License.

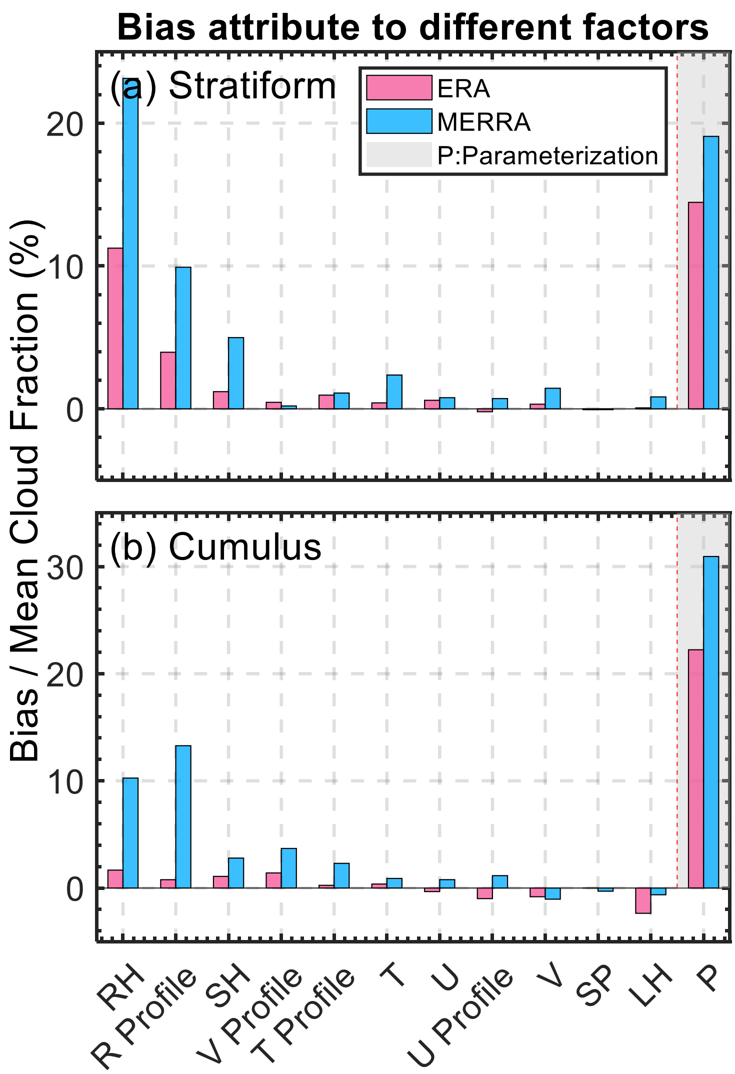


Figure 12: Attribution of bias between observed and reanalysis on cloud fractions to various meteorological factors and parameterization schemes for stratiform (a) and cumulus (b) cloud scenarios. The bars represent the normalized bias (bias divide mean cloud fraction) contributed by each factor: relative humidity profile (RH), meridional wind profile (V Profile), temperature profile (T Profile), zonal wind profile (U Profile), surface pressure (SP), latent heat flux (LH), and parameterization (P). All profiles took on morning (06:00 LT). Light blue bars indicate biases identified in the ERA reanalysis dataset, while pink bars represent biases in the MERRA reanalysis dataset. The dashed red line marked 'P' denotes biases attributed specifically to the parameterization within the reanalysis models. This analysis uses the DNN to discern the impact of each factor (ranked from highest to lowest) on the discrepancy in cloud fraction estimates between observations and reanalysis models.