

Machine-Learned Cloud Classes From Satellite Data For Process-Oriented Climate Model Evaluation

Polagangu Venkata Sowmya¹, Mr. M. Malakondrayudu²

PG scholar, Department of Computer Science & Engineering, Prakasam Engineering College, Kandukur.

Associate Professor, Department of Computer Science & Engineering, Prakasam Engineering College, Kandukur.

Abstract

Although clouds are crucial in controlling climate change, they are challenging to replicate in Earth system models (ESMs). Enhancing cloud representation is a crucial step towards more reliable climate change forecasts. In order to enhance comprehension of cloud representation and associated processes in climate models, this work presents a novel machine-learning framework based on satellite data. Coarse data may be assigned distributions of known cloud kinds using the suggested technique. It enhances the consistency of cloud process analysis and makes it easier to evaluate clouds in ESMs more objectively. Using cloud type labels from Cloud Sat as ground truth, the technique is based on deep neural networks labeling satellite data from the MODIS instrument with cloud categories established by the World Meteorological Organization (WMO). The technique works with datasets that provide physical cloud variable information at a temporal resolution that is high enough to be equivalent to MODIS satellite data. We use the technique using alternative satellite data, coarse-grained to usual resolutions of climate models, from the Cloud_cci project (ESA Climate Change Initiative). Our technique works with the common horizontal resolutions of ESMs, and the resultant cloud type distributions are physically consistent. We suggest that important variables needed by our approach be produced for next analysis of ESM data. This will make it possible to assess clouds in climate models more methodically by using tagged satellite data.

Index Terms— Climate modeling, clouds, Cloud Sat, CUMULO dataset, ESA Cloud_cci, machine learning, Moderate Resolution Imaging Spectroradiometer (MODIS), process-oriented model evaluation.

1. Introduction

EARTH system models (ESMs, often called climate models) are valuable resources for both projecting climate change under many likely future scenarios and enhancing our knowledge of the current climate. Nonetheless, a significant obstacle for ESMs continues to be simulating clouds and how they interact with the climate system [1]. One of the main causes of inter-model spread has been found to be the way clouds are represented in these models [2], [3]. Therefore, resolving these challenges requires an enhanced cloud process representation in ESMs [4]–[6]. Long-term satellite products with near-global coverage provide observations that are often used to evaluate model performance [e.g. 7], [8]. These data have shown to be highly suitable for the assessment of climate models. However, the constraints and uncertainties of the observational products themselves—such as biases or inconsistent geographical and temporal coverage—limit this typical method in part [9]. We provide a novel method for evaluating ESMs that aims to alleviate some of the perceived drawbacks of using traditional observational data while also making process-oriented cloud assessment in climate models easier. We make advantage of preexisting information about the features of various cloud classes, which are derived from the World Meteorological Organization's (WMO) taxonomy of cloud types. Utilizing this earlier information, cloud operations may be emphasized for further analysis. Our method applies machine learning-based cloud

categorization techniques for satellite data [10]–[15] to climate models, which is a new discovery. Although machine learning-based cloud categorization is not a novel concept [e.g. 16], it has only recently been practical for large-scale applications because of the rise in processing power that is now accessible and the fact that the various approaches have varied characteristics. The supervised vs unsupervised nature of categorization techniques is a key differentiator. Whereas the latter seeks to automatically discover unique new classes, the former depends on already given classes. While unsupervised approaches provide the user more flexibility over the composition of the classes, supervised classification makes the assumption that the classes allocated to them are appropriate for the task at hand. Consequently, supervised approaches need a collection of labeled data but enable interpretation of the final findings without the need for extra analytic stages [11], [14]. Unsupervised approaches are preferred if finding as different classes as feasible is the aim or if there are no accessible previously labeled data [12], [13]. As far as we are aware, no high-resolution (0.1 km) cloud class-labeled satellite data have been used so far for ESM analysis and assessment. Compared with rather coarse classifications, such as those employed, for instance, in the D-Series of the International Cloud Climatology Project (ISCCP, [17]), labeled datasets allow for a more comprehensive and direct interpretation of cloud types in the corresponding satellite data. The output of satellite simulators from models and clustering for satellite products have been employed in previous works for unsupervised cloud categorization [6, 18, 19, 19]. The selected clusters in these investigations are then allocated morphological cloud regimes based on the average physical attributes of each cluster. A categorization like this provides important insights into how different models more precisely depict clouds than would be possible with a basic climatology of physical factors. Nevertheless, the findings may be impacted by uncertainties and artifacts generated by the satellite simulators, in addition to being based on the relatively low resolution of 1280 km² of the ISCCP-D1 [17] product [20], [21]. In a recent research, the quantity of each of the four cloud classifications per cell was assigned using a convolution neural network on 14000 km² grid cells [22]. The classes in [22] came from WMO classes found from surface measurements, and the technique works with the output of climate models. In order to study certain cloud features like precipitation or radioactive impacts, several research have categorized satellite data by cloud regime [23], [24]. Using the 1° × 1° resolution ISCCP-H output, cloud regime clustering techniques have recently been applied to current-generation climate models [25].

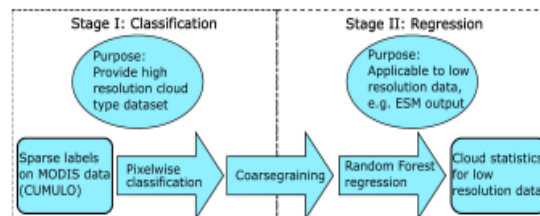


Fig.1. Two stages of machine-learning—a classifier and a regression model—are required to obtain cloud-type predictions on datasets with low horizontal resolution.

2. Literature Survey

P. P. Vignesh, J. H. Jiang, P. Kishore, H. Su, T. Smay, N. Brighton, and I. Velicogna,[1] In terms of skills and multimodal agreement, the seasonal and regional fluctuations of cloud fractions are compared across two generations of global climate model ensembles, namely the Coupled Model Intercomparison Project-5 (CMIP5) and CMIP6, across the historical period. Compared to what the CMIP5 model predicted, we find a larger dispersion of historical cloud percentage changes in the CMIP6. In comparison to CMIP5, the worldwide mean cloud percentages in CMIP6 rose by around 4.5%, which was attributable to more changes in the northern than in the southern hemisphere. To comprehend the cloud fraction uncertainties in CMIP6 models, the CALIPSO_CLOUSAT data are used to verify the CMIP6 cloud fractions from recent years. At lower altitudes, there is a mean difference of 0.5% between the CMIP6 ensemble mean of cloud percentages and the data, indicating good agreement. In the upper troposphere, the CMIP6 cloud percentages are greater than the data at higher latitudes in both hemispheres, and the biases differ amongst models. The model has a 3% larger bias across the tropics, further revealing the spatial mismatch between the ensemble and data. Additionally, utilizing

estimations of cloud fraction trends based on the robust regression approach, we detected a significant trend that has been occurring in the northern hemisphere since the mid-20th century. Finally, we use a straightforward regression approach to minimize the discrepancies between the model and data. The root mean square value dropped by over 28% and the correlation significantly enhanced, demonstrating how well the model and modified data compare.

L. Denby, [2] The main source of ambiguity in estimations of climate sensitivity is the way shallow trade wind convective clouds are represented in climate models. Specifically, little is known about the radiative effect of cloud spatial structure. The first unsupervised neural network model that can identify cloud structure regimes in satellite photos on its own is presented in this paper. Equipped with 10,000 GOES-16 satellite photos (spanning the tropical Atlantic and boreal winter), the discovered regimes exhibit a hierarchical structure of organizational sizes, whereby sub-clusters possess unique radiative characteristics. The methodology enables the objective analysis of extremely large data sets by eliminating the need for laborious and subjective hand-labeled data based on predetermined structures. With the help of cloud formations that emerge in both, the model allows for the objective comparison of model behavior with observations, as well as the study of environmental circumstances in various organizational regimes and during regime transitions. These capabilities make it possible to identify physical linkages in cloud processes that were previously unknown, which improves the depiction of clouds in weather and climate models.

W. J. Marais, R. E. Holz, J. S. Reid, and R. M. Willett [3] Multichannel spectral tests on individual pixels (i.e., fields of view) are used in the current cloud and aerosol detection techniques for multispectral radiometers, such as the Moderate Resolution Imaging Spectroradiometer (MODIS) and Visible Infrared Imaging Radiometer Suite (VIIRS). Phase and cloud top height are two examples of statistical factors that are often used in cloud and aerosol algorithms to classify clouds. Multispectral microphysical retrievals, on the other hand, give cloud classification information. Since clouds and aerosols have similar spectral characteristics in coarse-spectral-resolution studies, there is ambiguity in distinguishing optically thick aerosols using these approaches. Furthermore, since low-altitude cloud regimes have similar spectral characteristics, it is challenging to determine cloud regimes (such as stratiform and cumuliform) from only spectral observations. Deep neural network-based improvements in computer vision have opened up new possibilities for maximizing the coherent spatial information found in multispectral pictures. We exhibit advances in the ability to distinguish between cloud and severe aerosols, as well as an increased capacity to categorize different kinds of clouds, using a mix of machine learning approaches and a novel methodology to generate the required training data. An optimized version of the NASA Worldview platform, which offers a user-friendly interface for compiling a human-labeled database of cloud and aerosol types, was used to construct the labeled training dataset. Using MODIS cloud and aerosol data and independent Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), the convolutional neural network's (CNN) accuracy in classifying aerosols and cloud types was measured.

Arthur Grundner, Tom Beucler, Pierre Gentine, Fernando Iglesias-Suarez [4] Using deep learning with training data from storm-resolving model (SRM) simulations is a potential way to enhance cloud parameterizations inside climate models and therefore climate forecasts. The ICOSahedral Non-hydrostatic (ICON) modeling framework is a perfect target to create neural network (NN) based parameterizations for sub-grid scale processes, since it allows simulations ranging from numerical weather prediction to climate forecasts. We use coarse-grained data from realistic regional and global ICON SRM simulations to train neural network (NN) based cloud cover parameterizations inside the ICON framework. We configure three kinds of NNs for diagnosing cloud cover using coarse-grained atmospheric state data, which vary in the degree of vertical locality they assume. Using coarse-grained data with comparable geographical properties to their training data, the NNs correctly predict sub-grid size cloud cover. Furthermore, the sub-grid scale cloud cover of the regional SRM simulation may be replicated by globally trained NNs. We identify the cause for our column-based NN's inability to completely generalize from the global to the regional coarse-grained SRM data as an overemphasis on particular humidity and cloud ice, using the game-theory based interpretability library SHapley Additive exPlanations. Additionally, the interpretability tool shows a local correlation between the thermodynamic environment and cloud cover forecasts made by locally and globally trained column-based NNs, as well as

similarities and variations in feature relevance between them. Our findings demonstrate the capability of deep learning to extract interpretable and accurate cloud cover parameterizations from global SRMs, and they imply that neighborhood-based models could provide a reasonable intermediate ground between generalizability and accuracy.

3. Existing System

One of the main sources of uncertainty in future climate estimates is cloud feedback uncertainty in climate models. Thus, in order to guarantee the correctness of climate models, cloud simulation assessment and development are crucial. With regard to global mean near-surface temperature (GMST), we examine cloud biases and cloud change in climate models in relation to satellite data. We next connect these findings to equilibrium climate sensitivity, transient climate response, and cloud feedback. In order to achieve this, we create a supervised deep convolutional artificial neural network that can identify different types of clouds based on low-resolution ($2.5^\circ \times 2.5^\circ$) daily mean top-of-atmosphere radiation fields for shortwave and longwave wavelengths. These radiation fields correspond to the cloud genera recorded by human observers in the Global Telecommunication System (GTS) and are recognized by the World Meteorological Organization (WMO). We apply this network to the output of the Coupled Model Intercomparison Project Phases 5 and 6 (CMIP5 and CMIP6), the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis version 5 (ERA5), and the Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA-2) reanalyses. We train the network using top-of-atmosphere radiation retrieved by the Clouds and the Earth's Radiant Energy System (CERES) and GTS. We examine the differences in cloud types between satellite observations and simulations. We connect biases to climate sensitivity and find a negative linear association between model equilibrium climate sensitivity (ECS), transient climate response (TCR), and cloud feedback, and the root mean square error of cloud type occurrence obtained from the neural network. The model ensemble's statistical connection favors models with stronger cloud feedback, TCR, and ECS. This association, however, may be the result of the ensemble's very modest size or the decoupling of predicted cloud change in the future from current biases. With the use of the abrupt-4 \times CO₂ CMIP5 and CMIP6 experiments, we demonstrate that models that simulate stratiform clouds that are decreasing and those that are increasing typically have higher ECS than models that simulate stratiform clouds that are increasing and that could also partially explain the relationship between the model ECS and the model cloud type occurrence error.

3.2 Proposed System

We provide a novel method for evaluating ESMs that aims to alleviate some of the perceived drawbacks of using traditional observational data while also making process-oriented cloud assessment in climate models easier. We make advantage of preexisting information about the features of various cloud classes, which are derived from the World Meteorological Organization's (WMO) taxonomy of cloud types. Utilizing this earlier information, cloud operations may be emphasized for further analysis. Our strategy applies machine learning-based cloud categorization techniques recently developed for satellite data to climate models. Although machine learning-based cloud categorization is not a novel concept [e.g. 16], it has only recently been practical for large-scale applications because of the rise in processing power that is now accessible and the fact that the various approaches have varied characteristics. The supervised vs unsupervised nature of categorization techniques is a key differentiator. Whereas the latter seeks to automatically discover unique new classes, the former depends on already given classes. While unsupervised approaches provide the user more flexibility over the composition of the classes, supervised classification makes the assumption that the classes allocated to them are appropriate for the task at hand. As a consequence, supervised approaches need a set of labeled data yet enable interpretation of the final findings without the need for extra analytic stages. Unsupervised approaches are better if finding as different classes as feasible is the aim, or if there are no accessible previously tagged data.

4. List Of Modules

- Service Provider
- View and Authorize Users

- Remote User

4.1 MODULE DESCRIPTION

4.1.1 Service Provider

The Service Provider must provide a valid user name and password to log in to this module. He can do some tasks after logging in successfully, such Train & Test Data Sets, View Accuracy of Datasets in Bar Chart, View Accuracy of Datasets in Results View the Climate Class Type Prediction, View the Climate Ratio Prediction, Download the Predicted Data Sets, View All Remote Users and View Cloud Class for climate Type Ratio Results.

4.1.2 View and Authorize Users

The administrator may see a list of all enrolled users in this module. In this, the administrator may see user information such name, email address, and address, and they can also approve people.

4.1.3 Remote User

There are n numbers of users present in this module. Prior to beginning any actions, the user must register. The user's information is saved in the database when they register. Upon successful registration, he must use his permitted user name and password to log in. After logging in successfully, the user may do several tasks such as creating an account, predicting a cloud class based on climate type, and seeing their profile.

5. Architecture Diagram

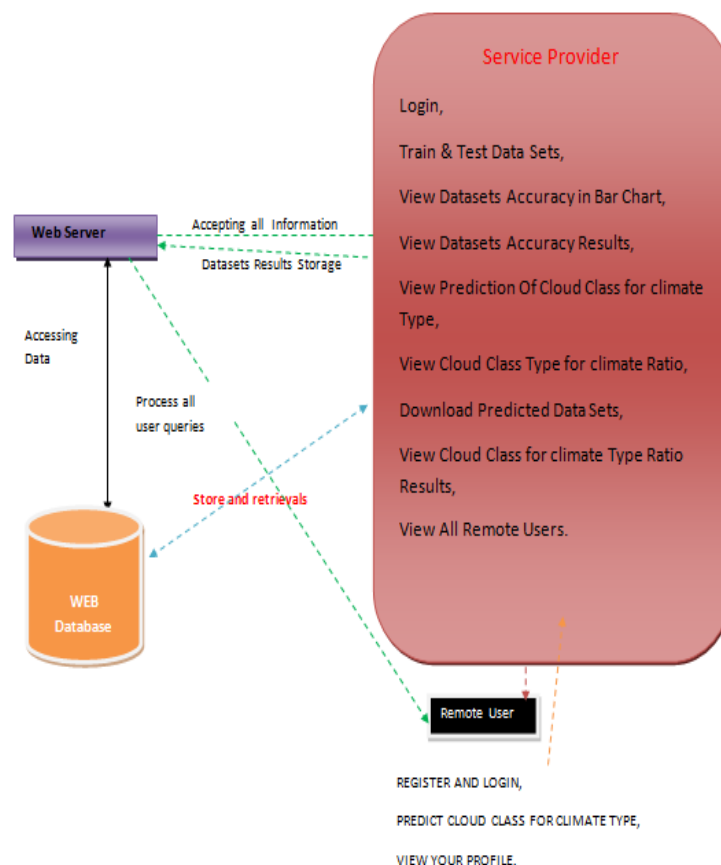


Fig. 5.1 Architecture Diagram



Fig.6.5. this is the Admin login page

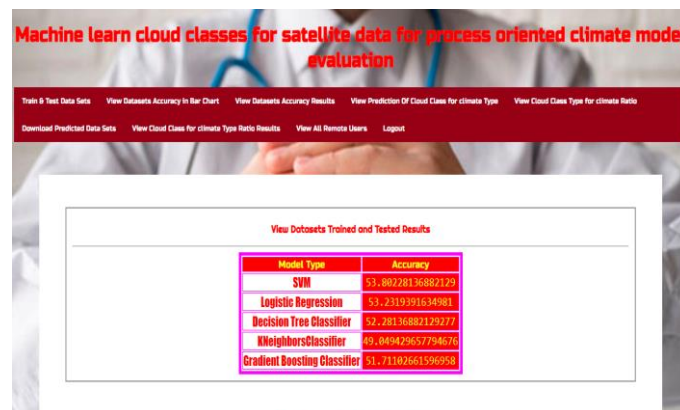


Fig.6.6. This is the admin page. These are the algorithms we have used in this project

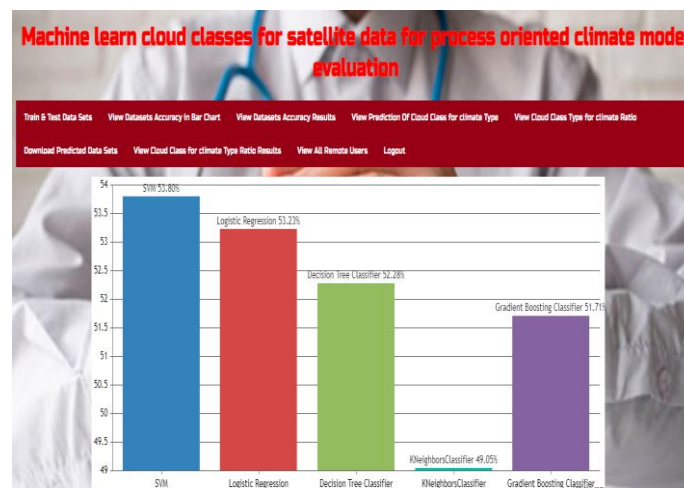


Fig.6.7. the algorithms are represented in bar chat

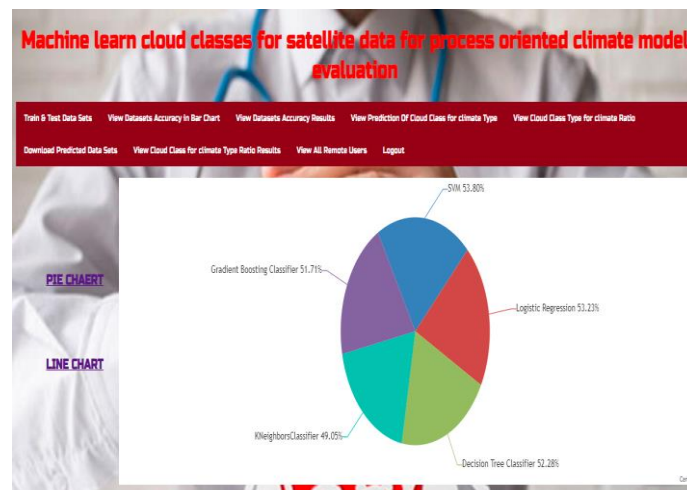


Fig.6.8. the algorithms are represented in pie chat



Fig.6.9. In the dataset we have two output lables that is partly cloud and light rain that is represented in percentage



Fig. 6.10.The data represented in pie chart

7. Conclusion and Discussion

Testing the regression model on data that hasn't been observed yet reveals temporal and physical consistency in the findings across all studies. This is the method's main objective, which is to assess physical processes. Therefore, even if the classification results do not precisely replicate the label distribution in the original data, we may be certain that the findings are meaningful. The classifier's predictions about the quantity of St may be partially accounted for by the short number of training samples and the physical similarities between Sc and St. We find that for datasets with a horizontal resolution common for climate model sizes, pixel-wise labeled data are thus viable as a foundation for training a regression model that learns cloud class distributions. Our findings generally imply that the approach is appropriate for a process-oriented evaluation of clouds that climate models generate. This may be done inside the cloud classes using the expected distributions, which has several benefits. First, the findings may be analyzed in terms of cloud classes that are precisely specified by the underlying classification system, eliminating a layer of subjective interpretation (Cloud Sat, [28]). The essential processes driving formation and development vary between the cloud classes, thus as we are utilizing the widely recognized WMO classes, the resultant cloud class distributions may then be examined and understood in a process-based way. This significantly streamlines the examination and assessment of certain physical attributes associated with cloud dynamics inside climate models. Second, the climate models are implicitly super-resolved horizontally while the deep learning algorithm learns from high-resolution 3-dimensional data. This analysis considers information about the vertical structure of clouds, meaning that learning from a combination of 2- and 3-dimensional data may benefit from vertical information not included in the cloud top view. The ability of the approach to resolve phenomena on seasonal and regional sizes makes it possible to pinpoint spatiotemporal regions where clouds are inaccurately represented. This might be carried out, for instance, to study the horizontal extent, dependency on feature values, and temporal development of the low level clouds present in the subtropics west of the continents. However, because of the multi-stage process's design, there are some restrictions. For example, it is challenging to produce accurate predictions on particular grid cells since the regression was built using 2-dimensional, spatially averaged source data. As a consequence, the expected cloud percentage varies by a factor of two or more for a number of samples. Furthermore, this approach may yet be improved, as seen by the underrepresentation of the Cu class and the restricted precision of the St class. Given that the Cloud Sat algorithm struggles to discern between St and Sc, this most likely results from the Cloud Sat ground truth itself, at least in part. Therefore, when using these or related methods, we advise combining these classes. Noisy satellite retrievals may magnify or obscure certain elements of the expected cloud distribution, such as the large proportions of Ns around the Antarctic shore. Clouds may be difficult to define using passive sensors such as MODIS, especially at high latitudes. Since our ML models are trained on immediate observations, they do not perform well when applied to temporally averaged data. For the pixel-wise categorization, using geostationary data (available, for example, every 30 minutes; GOES satellite, [37]) rather than MODIS data (available, for example, just twice a day) may provide better results. Other atmospheric factors including convection and rainfall have also been studied using this methodology [38], [39]. Because of the high and constant temporal resolution of the data, the physical features of the anticipated clouds could then be safely averaged across time, enabling the regression model to train on data that was more like to the output of conventional ESMs. Nevertheless, our method cannot resolve the processes at large temporal scales that need to be assessed. When the RF is trained on temporally averaged data, this will continue to be a problem and contribute to the poor regression performance for monthly mean data. This in turn suggests that, as opposed to utilizing climatologically means from long-term simulations, this technique is ideal to uncover problems in the model rather rapidly. This is because, with model output available for less than a year, we would anticipate an erroneous depiction of the global and regional cloud distributions to be visible now.

References

- [1] P. P. Vignesh, J. H. Jiang, P. Kishore, H. Su, T. Smay, N. Brighton, and I. Velicogna, "Assessment of CMIP6 Cloud Fraction and Comparison with Satellite Observations,"

-
- Earth and Space Science, vol. 7, no. 2, pp. 1–21, 2020.
- [2] J.-L. Dufresne and S. Bony, “An assessment of the primary sources of spread of global warming estimates from coupled atmosphere–ocean models,” *Journal of Climate*, vol. 21, no. 19, pp. 5135–5144, 2008.
 - [3] M. D. Zelinka, T. A. Myers, D. T. McCoy, S. Po-Chedley, P. M. Caldwell, P. Ceppi, S. A. Klein, and K. E. Taylor, “Causes of Higher Climate Sensitivity in CMIP6 Models,” *Geophysical Research Letters*, vol. 47, no. 1, pp. 1–12, 2020.
 - [4] S. Bony, B. Stevens, D. M. Frierson, C. Jakob, M. Kageyama, R. Pincus, T. G. Shepherd, S. C. Sherwood, A. P. Siebesma, A. H. Sobel, M. Watanabe, and M. J. Webb, “Clouds, circulation and climate sensitivity,” *Nature Geoscience*, vol. 8, no. 4, pp. 261–268, 2015.
 - [5] T. Schneider, J. Teixeira, C. S. Bretherton, F. Brient, K. G. Pressel, C. Schär, and A. P. Siebesma, “Climate goals and computing the future of clouds,” *Nature Publishing Group*, vol. 7, no. 1, pp. 3–5, 2017.
 - [6] K. D. Williams and M. J. Webb, “A quantitative performance assessment of cloud regimes in climate models,” *Climate Dynamics*, vol. 33, no. 1, pp. 141–157, 2009.
 - [7] A. Lauer, V. Eyring, M. Righi, M. Buchwitz, P. Defourny, M. Evaldsson, P. Friedlingstein, R. de Jeu, G. de Leeuw, A. Loew, C. Merchant, B. Müller, T. Popp, M. Reuter, S. Sandven, D. Senfleben, M. Stengel, M. V. Roozendaal, S. Wenzel, and U. Willen, “Benchmarking CMIP5 models with a subset of ESA CCI Phase 2 data using the ESMValTool,” *Remote Sensing of Environment*, vol. 203, pp. 9–39, Dec. 2017.
 - [8] H. Kawai, S. Yukimoto, T. Koshiro, N. Oshima, T. Tanaka, H. Yoshimura, and R. Nagasawa, “Significant improvement of cloud representation in the global climate model mri-esm2,” *Geoscientific Model Development*, vol. 12, no. 7, pp. 2875–2897, 2019.
 - [9] C. Jakob, “An improved strategy for the evaluation of cloud parameterizations in gcms,” *Bulletin of the American Meteorological Society*, vol. 84, no. 10, pp. 1387–1402, 2003.
 - [10] S. Rasp, H. Schulz, B. B. Stevens, and S. Bony, “Combining crowd-sourcing and deep learning to understand meso-scale organization of shallow convection,” in *AGU Fall Meeting 2019*, AGU, 2019.
 - [11] V. Zantedeschi, F. Falasca, A. Douglas, R. Strange, M. Kusner, and D. Watson-Parris, “Cumulo: A dataset for learning cloud classes,” in *Proceedings of the NeurIPS 2019 Workshop Tackling Climate Change with Machine Learning*, Climate Change AI, 2019.
 - [12] L. Denby, “Discovering the Importance of Mesoscale Cloud Organization Through Unsupervised Classification,” *Geophysical Research Letters*, vol. 47, no. 1, pp. 1–10, 2020.
 - [13] T. Kurihana, E. Moyer, R. Willett, D. Gilton, and I. Foster, “Data-Driven Cloud Clustering via a Rotationally Invariant Autoencoder,” *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–25, 2021. arXiv: [2103.04885](https://arxiv.org/abs/2103.04885).
 - [14] C. Zhang, X. Zhuge, and F. Yu, “Development of a high spatiotemporal resolution cloud-type classification approach using Himawari-8 and CloudSat,” *International Journal of Remote Sensing*, vol. 40, no. 16, pp. 6464–6481, 2019.
 - [15] W. J. Marais, R. E. Holz, J. S. Reid, and R. M. Willett, “Leveraging spatial textures, through machine learning, to identify aerosols and distinct cloud types from multispectral observations,” *Atmospheric Measurement Techniques*, vol. 13, no. 10, pp. 5459–5480, 2020.
 - [16] J. Lee, R. C. Weger, S. K. Sengupta, and R. M. Welch, “A Neural Network Approach to Cloud Classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 5, pp. 846–855, 1990.
 - [17] W. B. Rossow and R. A. Schiffer, “Advances in Understanding Clouds from ISCCP,” *Bulletin of the American Meteorological Society*, vol. 80, no. 11, pp. 2261–2287, 1999.

- [18] C. Jakob and G. Tselioudis, "Objective identification of cloud regimes in the Tropical Western Pacific," *Geophysical Research Letters*, vol. 30, no. 21, pp. 1–4, 2003.
- [19] N. D. Gordon, J. R. Norris, C. P. Weaver, and S. A. Klein, "Cluster analysis of cloud regimes and characteristics dynamics of midlatitude synoptic systems in observations and a model," *Journal of Geophysical Research D: Atmospheres*, vol. 110, no. 15, pp. 1–13, 2005.
- [20] R. Pincus, S. Platnick, S. A. Ackerman, R. S. Hemler, and R. J. Patrick Hofmann, "Reconciling simulated and observed views of clouds: MODIS, ISCCP, and the limits of instrument simulators," *Journal of Climate*, vol. 25, no. 13, pp. 4699–4720, 2012.