ADVANCING EARTH AND SPACE SCIENCE

JAMES | Journal of Advances in Modeling Earth Systems®

RESEARCH ARTICLE

10.1029/2021MS002625

Key Points:

- Neural network ensembles are able to detect how different meteorological factors are between E3SM, E3SM-FIVE, and observations
- E3SM-FIVE high vertical resolution simulation improves marine warm low clouds fractions given the same meteorological factors
- Meteorological factors between the three data sets are large enough to dominate the differences in marine warm low cloud fractions

Correspondence to:

Y.-S. Chen,
yaosheng.chen@noaa.gov

Citation:

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

Received 14 MAY 2021
Accepted 9 OCT 2021

Model Evaluation and Intercomparison of Marine Warm Low Cloud Fractions With Neural Network Ensembles

Yao-Sheng Chen[1,2], **Takanobu Yamaguchi**[1,2], **Peter A. Bogenschutz**[3], and **Graham Feingold**[2]

[1]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA, [2]NOAA Chemical Sciences Laboratory, Boulder, CO, USA, [3]Lawrence Livermore National Laboratory, Livermore, CA, USA

Abstract Low cloud fractions (LCFs) and meteorological factors (MFs) over an oceanic region containing multiple cloud regimes are examined for three data sets: one Energy Exascale Earth System Model (E3SM) simulation with the default 72-layer vertical grid (E3SM72), another one with 8-times vertical resolution via the Framework for Improvement by Vertical Enhancement (E3SM×8), and one with MFs from ERA5 reanalysis and LCFs from the CERES SSF product (ERA5-SSF). Neural networks (NNs) are trained to capture the relationship between MFs and LCF and to select the best-performing MF subsets for predicting LCF. NN ensembles are used to (a) confirm the performance of selected MF subsets, (b) to serve as proxy models for each data set to predict LCFs for MFs from all data sets, and (c) to classify MFs into those in shared and uniquely occupied MF subspaces. Overall, E3SM72 and E3SM×8 have large fractions of MFs in shared MF subspace, but less so near the Californian and Peruvian stratocumulus decks. E3SM×8 and ERA5 have small fractions of MFs in shared MF subspace but greater than E3SM72 and ERA5, especially in the Southeast Pacific. The differences in LCFs between three pairs of data sets are decomposed into those associated with the differences in the LCF-MF relationship and those involving different MFs. Given the same MFs, LCFs produced by E3SM×8 are greater than those produced by E3SM72 but are still different from those in ERA5-SSF. In general, the shift in MFs dominates the difference in the LCFs.

Plain Language Summary Marine warm low clouds are critical for both present day and future climate because they reflect a lot of solar energy back to space. To make more reliable projections of our changing climate, scientists need to improve these clouds in climate models. One question that scientists ask is, why do the climate models predict so much less marine warm low cloud cover than the satellites see? Is it because the models misrepresent the meteorology (like temperature and humidity) or because they are not able to produce enough clouds even if they predict the meteorology well. In this work, we use neural networks, a machine learning technique, to answer these questions. We find that our recent efforts to improve a climate model help the model produce more marine warm low clouds given the same meteorology; these efforts also lead to changes in the meteorology predicted by the model. Further model improvements are needed to bring the model predictions closer to the observations.

1. Introduction

Marine low clouds cover vast areas of the eastern subtropical oceans in the form of stratocumulus decks, which transition to shallow cumulus toward the west and lower latitudes. They are also found in the cold sector of midlatitude cyclones and in the polar regions, very often as mixed-phase clouds. By increasing the planetary albedo, marine low clouds play a critical role in Earth's energy budget (e.g., L'Ecuyer et al., 2019). Because they reside close to the Earth's surface, their radiative impacts are primarily in the shortwave and depend on their macrophysical (e.g., cloud fraction) and microphysical (e.g., drop size distribution) properties.

Previous studies have related the marine cloud properties (CPs) to a variety of meteorological factors (MFs)—quantities that characterize the marine low cloud environment. These MFs can be derived from observations or large-scale model data at various spatial and temporal scales (e.g., Bony et al., 2004; Kubar et al., 2011; Mieslinger et al., 2019; Myers & Norris, 2013; Weaver & Ramanathan, 1997). Several compound

MFs, that is, those derived from more basic thermodynamic and dynamic properties, have been proposed and have shown a good correlation with CPs. Some of these compound MFs are based on a physical understanding of lower-tropospheric atmospheric structure and processes. For example, the lower-tropospheric stability (LTS) is defined as the potential temperature difference between 700 mb and surface (Klein & Hartmann, 1993); both the estimated inversion strength (EIS, Wood & Bretherton, 2006) and the estimated low-level cloud fraction (ELF, Park & Shin, 2019) take into account the empirical thermodynamic structure of the lower-troposphere and the former can be interpreted as a measure of effective stability; the estimated cloud top entrainment index (ECTEI) considers the impacts of cloud top entrainment instability (Kawai et al., 2017). A compound MF can also be derived following a data-driven approach. Wall et al. (2017) applied principal component analysis to transform four MFs and used the leading principal component as a single MF to effectively indicate the transition between the cold and warm sectors of midlatitude cyclones in the Southern Ocean and thus the CPs in different parts of these cyclones.

Within the CP-MF framework, the CP in a CP-MF data set is conceptually built on two components: (a) the MFs and (b) the relationship between the CPs and the MFs. For a model-based data set, the former component is related to the model's ability to reproduce the meteorological regimes while the latter component is directly controlled by its physical parameterizations. In practice, one commonly used approach is to first group the data by either the CPs (into cloud regimes, e.g., Jakob & Tselioudis, 2003; Tselioudis et al., 2000) or the MFs (into meteorological regimes, e.g., Barton et al., 2012; Bony et al., 2004; Bennhold & Sherwood, 2008; Mechem et al., 2018; Medeiros & Stevens, 2011; Mülmenstädt et al., 2012; Norris & Weaver, 2001; Su et al., 2013), and then examine the frequency of occurrence of the groups as well as the statistics of CPs and MFs within each group.

Previous works used this approach for model evaluation and intercomparison and studied whether the discrepancy in CPs between model-based and observation-based data sets is dominated by the differences in the MFs or the differences in the CP-MF relationships in these data sets. Medeiros and Stevens (2011) reported that the large-scale environments are similar between the models and observations they evaluated, implying that the large differences in the tropical low cloud amounts between the models and the observations are rooted in the models' physical parameterizations. Su et al. (2013) drew similar conclusions regarding their cloud properties of interest. We comment on two aspects of these studies. First, the cloud properties of interest were examined as a function of only one or two MFs, which are insufficient to characterize the clouds' environment (Nam & Quaas, 2013). Second, there is no quantification of the similarity between the MFs from modeled or observed data sets. Although some other studies (e.g., Ma et al., 2014) also attributed model errors to the physical parameterizations using short-term hindcast simulations that constrained the large-scale environments, there is emerging evidence pointing to the importance of the MFs in driving the spatial errors of modeled low stratiform cloud decks (Brunke et al., 2019).

Our primary concern in this study is the marine warm low cloud fraction (LCF) simulated with the Energy Exascale Earth System Model (E3SM), an Earth system model developed by the U.S. Department of Energy (E3SM Project, 2018). Like many other climate models, E3SM underestimates the marine LCF for low latitudes and subtropical regions (Y. Zhang et al., 2019). Recently, Bogenschutz et al. (2021) reported that using high vertical resolution in the lower troposphere in E3SM increases the LCFs in the climatological subtropical stratocumulus regimes. Lee et al. (2021) implemented the Framework for Improvement by Vertical Enhancement (FIVE, Yamaguchi et al., 2017), a framework that enables the host model to compute selected processes at the high vertical resolution, into E3SM and achieved similar improvements at a much lower computational cost. In this study, we apply the LCF-MF framework to understand the sources of LCF improvements from E3SM to E3SM-FIVE and the sources of LCF errors between model simulations and reanalysis and satellite data. We use neural networks (NNs) as a nonlinear regression tool to capture the relationships between the LCF and multiple MFs for two model-based data sets and one observation-based data set and use the NN ensemble to associate the differences in the LCFs to the differences in the MFs, and the differences in the LCF-MF relationship. We will introduce our data sets and methods in detail in Section 2, followed by results in Section 3, and conclude after some discussion in Section 4.

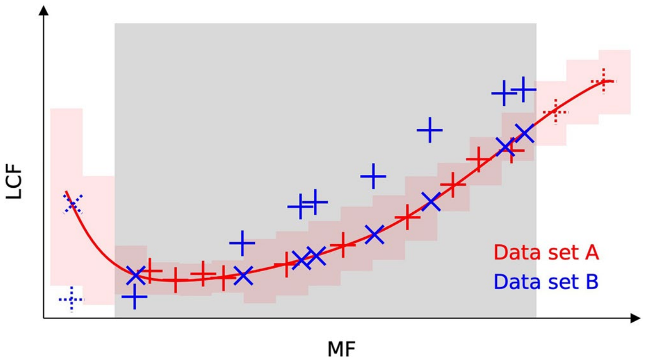

Figure 1. Scheme for key elements of the methods presented in this study. Red and blue "+"s represent the data points from two data sets: (A and B) Red curve represents the relation between low cloud fraction (LCF, denoted as f) and meteorological factor (MF, denoted as $\mathcal{M}$) captured by the mean of neural network (NN, denoted as $\mathcal{F}$) ensemble trained on the data from (a) Red shaded bars represent the standard deviation in LCF predicted by NN ensemble members. Blue "×"s represent the mean LCFs predicted by $\mathcal{F}_{\mathrm{A}}$ ensemble for $\mathcal{M}_{\mathrm{B}}$ (denoted as $\mathcal{F}_{\mathrm{A}}(\mathcal{M}_{\mathrm{B}})$). Symbols ("+"s and "×"s) in solid lines indicate data points with the MF falling in the MF subspace shared between the two data sets (denoted as $\mathcal{M}_{\mathrm{A},\cap\mathrm{B}}$ and $\mathcal{M}_{\mathrm{B},\cap\mathrm{A}}$); those in dotted lines indicate data points with MFs falling in the uniquely occupied MF subspace (denoted as $\mathcal{M}_{\mathrm{A},\cap\bar{\mathrm{B}}}$ and $\mathcal{M}_{\mathrm{B},\cap\bar{\mathrm{A}}}$). Gray shading indicates the region shared between $\mathcal{M}_{\mathrm{A}}$ and $\mathcal{M}_{\mathrm{B}}$.

2. Data and Methods

2.1. Conceptual Description of Methods

The idea of our approach is summarized in the schematic in Figure 1. Assume, for illustration purposes, that the LCF is determined by a single MF and that we have LCF and MF data from two data sets: A (represented by red "+"s) and B (blue "+"s). We first train an ensemble of NNs (the mean of which is represented by the red curve) to map the MF from A (denoted as $\mathcal{M}_{\mathrm{A}}$) to the LCF from the same data set (denoted as f_{A}, which is an abbreviation of $f_{\mathrm{A}}(\mathcal{M}_{\mathrm{A}})$ since the LCF from a data set is always associated with the MF from the same data set). By applying the NN ensemble (denoted as $\mathcal{F}_{\mathrm{A}}$) to $\mathcal{M}_{\mathrm{B}}$, we can predict what the LCF would be if $\mathcal{M}_{\mathrm{B}}$ were to occur in A (represented by blue "×"s, and we denote the prediction as $\mathcal{F}_{\mathrm{A}}(\mathcal{M}_{\mathrm{B}})$). The differences between $\mathcal{F}_{\mathrm{A}}(\mathcal{M}_{\mathrm{B}})$ (blue "×") and f_{B} given the same $\mathcal{M}_{\mathrm{B}}$ tell us how much difference in LCF between A and B is due to different LCF-MF relationships between the two data sets.

However, we caution that even a well-trained NN ensemble is only able to "interpolate"; it is not designed to "extrapolate" (note the upward-tilting part of the red curve in the region without data from A). So, not all blue "×"s represent reliable predictions. Inspired by Lohninger (1999) (see Lohninger, 2006), we expect the NN ensemble standard deviation (represented by the heights of red shaded bars) to be relatively small for regions densely populated by the training data, since the NN ensemble is more constrained there, but to become larger when the NN ensemble is applied to the MFs far away from the training data. (See Appendix A for a demonstration with a toy problem.). We can then classify $\mathcal{M}_{\mathrm{B}}$ into those residing in an MF subspace shared with A (i.e., sharing a range that is also occupied by some $\mathcal{M}_{\mathrm{A}}$, denoted as $\mathcal{M}_{\mathrm{B},\cap\mathrm{A}}$, blue symbols with solid line) and those falling into the MF subspace that is uniquely occupied by B (denoted as $\mathcal{M}_{\mathrm{B},\cap\bar{\mathrm{A}}}$, blue symbols with dashed line) based on $\mathcal{F}_{\mathrm{A}}(\mathcal{M}_{\mathrm{B}})$ ensemble standard deviation. Similarly, we can use $\mathcal{F}_{\mathrm{B}}$ to classify $\mathcal{M}_{\mathrm{A}}$. (We borrow the notation used in set theory, where $\mathrm{A} \cap \mathrm{B}$ means set members that belong to both A and B and $\mathrm{A} \cap \bar{\mathrm{B}}$ means set members that belong to A but not B. We add a comma after the name of the first data set in this notation to emphasize that whether the MF of a data point from this data set falls into the shared or uniquely occupied MF subspace is determined by the NN ensemble trained on the second data set listed in this notation. In other words, "A, ∩ B" is not interchangeable with "B, ∩ A").

Finally we can reliably calculate the difference between $\mathcal{F}_{\mathrm{A}}(\mathcal{M}_{\mathrm{B},\cap\mathrm{A}})$ and f_{B} associated with $\mathcal{M}_{\mathrm{B},\cap\mathrm{A}}$ (denoted as $f_{\mathrm{B},\cap\mathrm{A}}$). With this difference removed from the total difference between f_{A} and f_{B}, the remaining difference is due to the difference in MFs, including the different frequency distributions between $\mathcal{M}_{\mathrm{A},\cap\mathrm{B}}$ and $\mathcal{M}_{\mathrm{B},\cap\mathrm{A}}$ within the shared MF subspace as well as the differences between $\mathcal{M}_{\mathrm{A},\cap\bar{\mathrm{B}}}$ and $\mathcal{M}_{\mathrm{B},\cap\bar{\mathrm{A}}}$.

2.2. Data Sets

We investigate the monthly mean LCFs and MFs from three data sets. The observation-based data set comprises LCFs from CERES Edition 4A (Minnis et al., 2020) Single Scanner Footprint (SSF) product from 2003 to 2018 and MFs from ERA5 reanalysis (Hersbach et al., 2020) over the same time period. The SSF LCF, representing the cloudiness between the surface and 700 mb, is retrieved following Minnis et al. (2020) and based on the cloud masks determined by comparing the radiances observed by MODIS with estimated clear-sky radiances through a set of tests (Trepte et al., 2019). We use the average of monthly daytime and nighttime mean LCFs for both Terra and Aqua from the Level 3 SSF 1° monthly product. The ERA5 reanalysis is obtained as hourly data at 0.25° × 0.25° resolution and processed to monthly data at 1° × 1°. We refer to this data set as "ERA5-SSF."

The two model-based data sets come from two 16-year global simulations with E3SM and E3SM-FIVE. The E3SM simulation is performed using the standard release of E3SMv1 (Golaz et al., 2019) on its default

72-layer vertical grid. The E3SM-FIVE simulation uses an 8-times resolution vertical grid from the second layer above the surface to about 700 mb for three physical parameterizations: Cloud Layers Unified By Binormals (CLUBB, Golaz et al., 2002; Larson & Golaz, 2005; Larson, 2017), version two of the Morrison and Gettelman microphysical scheme (MG2, Gettelman et al., 2015), and the Rapid Radiative Transfer Model for GCMs (RRTMG, Iacono et al., 2000; Mlawer et al., 1997). It also calculates the vertical advection by the resolved flow on the same high resolution vertical grid to balance the parameterized entrainment due to turbulence. Other than this, the two simulations are configured and used in identical ways, with no adjustments of the model tunable parameters. Both simulations are driven by climatologically prescribed sea surface temperatures (SSTs) and forcings. The atmospheric component is run on an ne30np4 cubed-sphere horizontal grid, corresponding to a horizontal resolution of about 1°. The first year of each simulation is considered as spin-up and the monthly mean MFs and LCFs from the subsequent 15 years are used. Hereafter we refer to the two simulations and the data sets based on their results as “E3SM72” and “E3SM×8”.

Both E3SM and E3SM-FIVE diagnose the LCF as follows: the model first calculates three cloud fractions per grid box: (a) the deep convection cloud fraction from the deep convective mass flux predicted by the Zhang-McFarlane (ZM) deep convection scheme (G. J. Zhang & McFarlane, 1995), (b) the liquid cloud fraction by CLUBB from the predicted joint distributions of thermodynamic variables, and (c) the ice stratiform cloud fraction that linearly changes with the ratio between the sum of ice mass mixing ratio and the vapor mixing ratio and the saturation vapor mixing ratio over ice (Gettelman et al., 2010; Slingo, 1987). Then a total cloud fraction is calculated as the sum of the deep convection cloud fraction and the greater value between the liquid and ice stratiform cloud fractions. Finally, the LCF, the focus of our work, is diagnosed from the total cloud fractions between surface and the 700 mb assuming maximum-random overlap among cloudy areas in different layers. The critical difference between the E3SM LCF and the SSF LCF is that the former is diagnosed purely from macroscopic quantities while the latter fundamentally relies on the interaction between hydrometeor particles (microphysics) and radiation. Satellite instrument simulators can partially address this discrepancy but not without their own issues (Pincus et al., 2012). For the current study, we focus on the first-order relationship between the MFs and the model-diagnosed LCF and avoid the extra complications, as well as benefits, associated with instrument simulators. We do take the average of only the hourly LCFs closest to four local times (i.e., 1:30, 10:30, 13:30, and 22:30) to match the overpass times of Terra and Aqua.

Although our focus is on marine warm (liquid only), low clouds, we avoid the tricky problem of identifying this cloud regime from the data by including all the data over the oceanic region between 40° S, 40° N, 70° W, and 167° W. This region encompasses the Californian and Peruvian stratocumulus decks and the shallow cumulus regimes; it also contains other cloud regimes, including deep convection in the tropics and extratropical cyclones (along the southern and northern edges). However, we only comment on the findings related to warm, low clouds. Figure 2 shows an overview of the multi-year mean LCFs from the three data sets in the region of interest. Throughout this work, we report LCF as decimals, instead of percentages. Compared to the SSF LCFs, the LCF patterns produced by E3SM72 and E3SM×8 show lower LCF in both the Californian and Peruvian stratocumulus decks. The Californian stratocumulus decks in E3SM72 and E3SM×8 sit to the south of 20° N and further away from the North America coast. In the Southeast Pacific, the low clouds in E3SM72 and E3SM×8 do not extend far enough to the west of the Peruvian stratocumulus decks. E3SM×8 produces greater LCF than E3SM72 at most locations and improves the agreement with the SSF LCFs. The Californian stratocumulus deck in E3SM×8 shows similar LCFs to E3SM72 to the south of 20° N but covers the area to the north of 20° N with much higher LCF. The Peruvian stratocumulus deck in E3SM×8 extends toward the south of its counterpart in E3SM72 with LCFs that can be as much as 0.24 higher than in E3SM72.

2.3. NNs and NN Ensembles

We use the densely connected feedforward NN to capture the LCF-MF relationships. It maps an input (column) vector $\vec{m}_0$ containing n_0 MFs as input variables (called features in the machine learning literature) to the corresponding LCF (denoted as f) via a series of operations:

$$f \sim g_{N+1}(g_N(\ldots(g_2(g_1(\vec{m}_0))))). \tag{1}$$

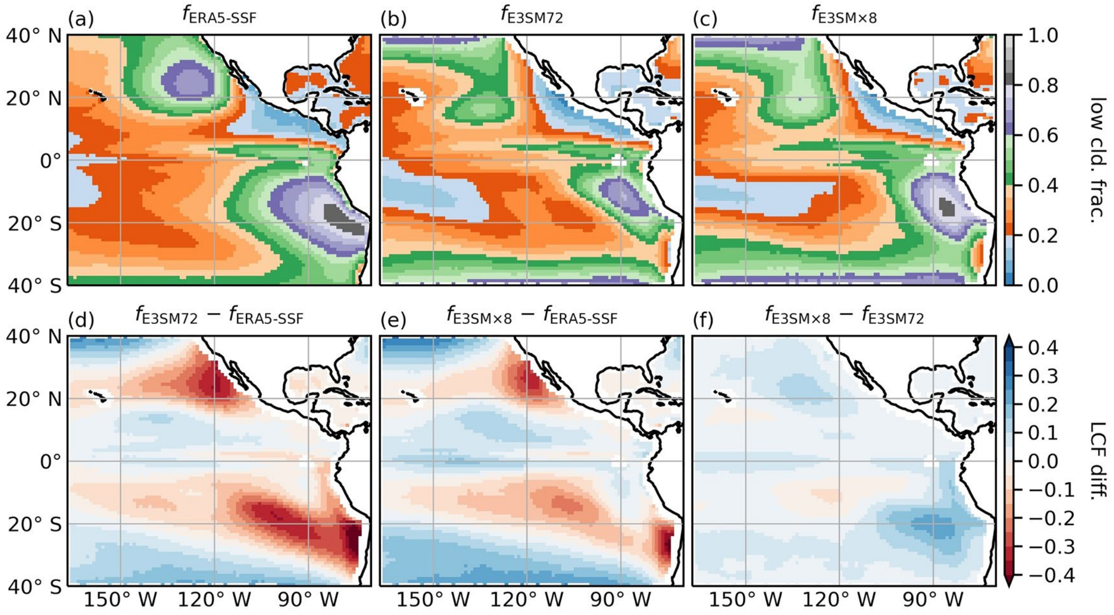

Figure 2. Mean low cloud fractions (LCFs or f) from SSF, E3SM72, and E3SM×8 (first row), and differences between them (second row).

The input vector $\vec{m}_0$ is first mapped to n_1 nodes in the first hidden layer, that is,

$$\vec{m}_1 = g_1(\vec{m}_0) = \phi_1(\mathbf{W}_1 \cdot \vec{m}_0 + \vec{b}_1), \tag{2}$$

where $\vec{m}_1$ is an $n_1 \times 1$ vector and $\mathbf{W}_1$, $\vec{b}_1$, and a nonlinear function ϕ_1 are the weight matrix, the bias vector, and the activation function for g_1. Similar operations continue until the nodes in the N^{th} hidden layers are mapped to f.

To begin the training of the NNs, the first task is to select the features. We start with an initial pool of 14 MFs that fall into four categories: quantities characterizing lower-tropospheric thermodynamics, that is, potential temperature (θ) and relative humidity (RH) at 1,000, 925, 850, and 700 mb; quantities describing atmospheric dynamics, that is, horizontal wind speed (U) at 1,000 and 700 mb and vertical air velocity in pressure coordinates (ω) at 500 mb; one column-integrated quantity, i.e., precipitable water vapor (PWV, also called total column water vapor); and surface sensible and latent heat fluxes (Q_h and Q_e). With these 14 initial MFs, we are essentially testing whether an NN can learn to diagnose the LCF for a variety of cloud regimes from coarse resolution soundings of θ, RH, U, and ω, supplemented with additional information on the atmospheric humidity (PWV) and the lower boundary conditions (e.g., surface fluxes) to the atmospheric columns under investigation. We do not include any compound MFs from the literature but train NNs to map each of the 16,383 subsets of these 14 basic MFs to the LCFs to search for the best-performing subsets. In other words, we train NNs for 14 1-MF subsets, 91 2-MF subsets, 364 3-MF subsets, all the way to the MF set containing all 14 MFs. The choice of analyzing monthly mean data at about $1° \times 1°$ itself is another feature selection decision we make. In other words, we exclude the MFs that are relevant at finer temporal and spatial scales.

The next step is to decide on the NN architecture, that is, the number of hidden layers N and the number of nodes in each hidden layer n_i (for $i = 1, 2, \ldots, N$). An NN's architecture controls its capacity. An NN with insufficient capability (too few hidden layers or nodes) is unable to memorize the pattern in the training data; an NN with excessive capability (too many hidden layers or nodes) tends to overfit the training data and does not generalize well to data that it has not seen before. With some trial and error, we choose one baseline architecture for each number of features (from 1 to 14) following two steps. We first search all NN architectures with five hidden layers and 24 to 32 nodes in each layer for those with 3,250 weights (i.e., 3,250 elements in all weight matrices combined). Then, from these NN architectures, we choose the one with the smallest standard deviation of the numbers of nodes in five hidden layers. For example, the baseline

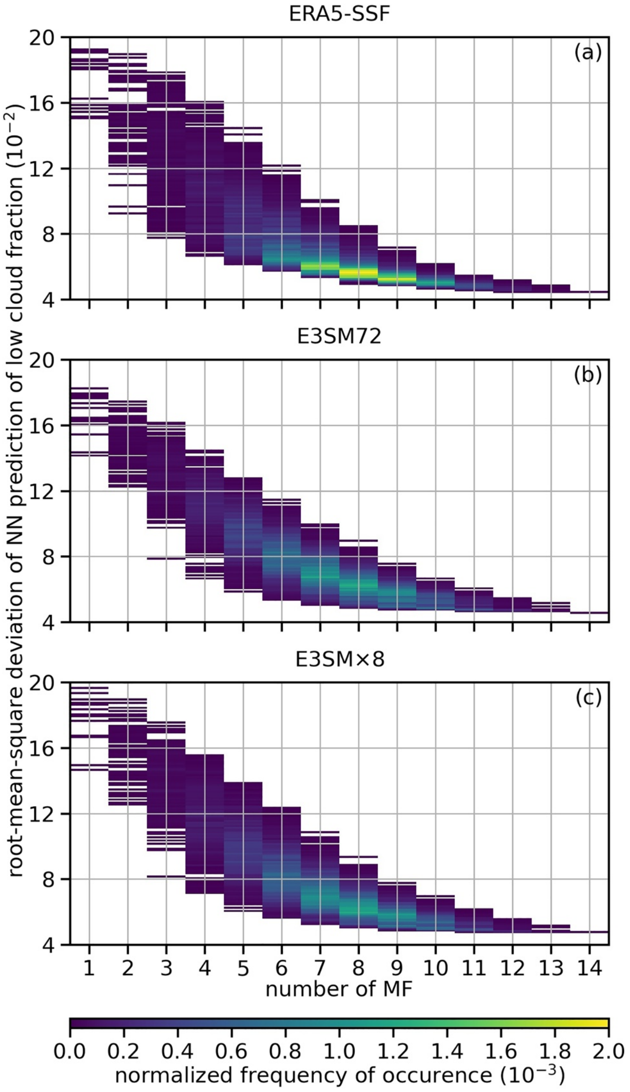

Figure 3. Distribution of performances of neural networks with baseline architectures and trained to map all 16,383 subsets of the 14 meteorological factors (MFs) to the low cloud fractions, sorted by the number of MFs in the MF subsets.

architecture for eight features has 28, 27, 27, 27, and 29 nodes in each layer from the input end to the output end, resulting in 3,389 parameters (i.e., 3,389 elements in all weight matrices and bias vectors combined).

We standardize all MFs for each data set by removing their means and scaling them to unit variance. The distributions of these MFs are all skewed, but we do not see noticeable improvements in the NN performances when we transform them to more symmetrical distributions. Therefore, no additional transformation is performed to the MFs. For the training, we linearly transform the LCF to between − 1 and 1 to match the choice of the activation functions used in the NN See below for details.) In pilot training, NN performances for E3SM-based data sets are significantly worse near the landmasses, including the North and South Americas, the Hawaii islands, and, to a lesser extent, the Galapagos Islands. We find that some oceanic grids near these landmasses in E3SM-based data sets have large surface geopotential, and the MFs at these grids are very noisy and contain more extreme values than at other grids. We exclude data points at grids with a surface height greater than 10 m to alleviate this issue. No data points are excluded for ERA5-SSF since the problem is less severe, probably because the noisy MFs near landmasses are smoothed out as we coarse-grain the original data. There are more than 1.2 million data points left in each data set. These data points are split into 68%:17%:15% for the training, validation, and test sets, respectively. On average, we train one parameter with at least 250 data points and only see a small degradation of the NN performance for the validation set relative to the performance for the training set, suggesting little overfitting.

All hidden layers of the NNs use the SWISH activation function (Ramachandran et al., 2017), which is simply

$$\phi(x) = \frac{x}{1 + e^{-\beta x}}, \tag{3}$$

with the parameter β set to 1; the output layer uses the tanh activation function. The NN parameters are initialized with the He uniform variance scaling initializer (He et al., 2015) and optimized with the ADAM optimizer (Kingma & Ba, 2014) for 100 epochs to minimize the mean squared error (MSE) in the NN predictions at a learning rate that starts at 10^{-3} and drops to 10^{-4} if the MSE for the validation data set plateaus. All NNs are constructed and trained with TensorFlow (Abadi et al., 2015), version 2.2.

Based on the results from NNs with baseline architectures, we perturb the NN architecture to generate NN ensembles for selected feature sets. For each feature set of interest, we again search all NN architectures with five hidden layers and 24 to 32 nodes in each layer but this time for NN architectures that satisfy one of two conditions: (a) number of weights equal to 3,249, 3,250, or 3,251 and (b) number of parameters within the baseline architecture's number of parameters ±1. For example, for eight features, the second condition identifies NN architectures with 3,388, 3,389, or 3,390 parameters. We then train NNs with these NN architectures to generate NN ensembles.

3. Results

3.1. Feature Selection and Performance of NNs

We first examine the performance of the NNs with baseline architectures, quantified by the root-mean-square deviation (RMSD) based on all data points in each data set, where the "deviation" is defined as the true LCF minus the prediction by the NNs. Figure 3 shows the distributions of the RMSDs for all 16,383

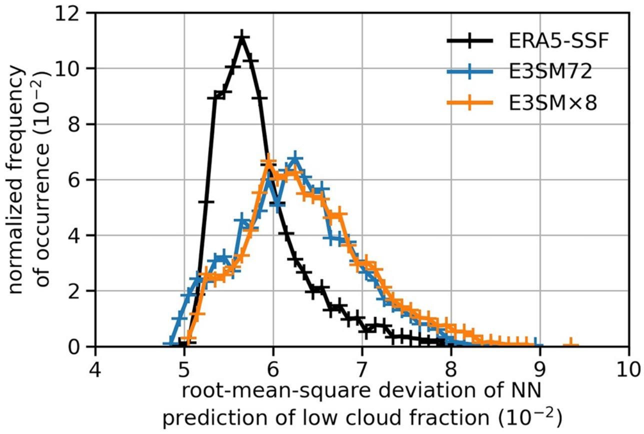

Figure 4. Distribution of performances of neural networks with the baseline architecture and trained to map all 3003 8-MF subsets of the 14 meteorological factors to the low cloud fractions.

NNs for each data set, sorted by the number of MFs. The overall performance and the best performance of the NNs improve as the number of MFs increases. The gain in the best performance is small beyond eight MFs. On the other hand, when NNs are trained on too few MFs (e.g., two MFs), the spatial patterns of the mean deviation show large magnitudes and clear correspondence with meteorological regimes, indicating insufficient information to characterize all meteorological regimes in our problem (e.g., See Figure 12). Based on these two findings, we focus on the NNs trained on 3003 8-MF subsets.

Figure 4 shows the distributions of the RMSDs for these 3003 NNs. The NNs for ERA5-SSF produce a RMSD distribution that peaks between 0.056 and 0.057. The best-performing 8-MF subsets have RMSDs that are slightly smaller than 0.050 but the worst-performing ones produce RMSDs as large as 0.084. The overall performances for the two E3SM-based data sets are worse than those for ERA5-SSF, but the smallest RMSDs are near 0.050, similar to those for ERA5-SSF. To provide some context for these values, note that if a predictor randomly samples a uniform distribution between 0 and 1 as its prediction for the LCF, the RMSD will fall between 0.36 and 0.38; if a predictor randomly draws from the true cloud fraction distribution as its prediction, the RMSD will fall between 0.25 and 0.29. Results depend on the LCF distribution for each data set.

We test the robustness of the rank of the subsets' performance determined by the NNs with the baseline architecture. We train NN ensembles using 341 NN architectures, identified following the aforementioned method, for the top six best-performing 8-MF subsets of each data set. Figure 5 displays the distributions of the RMSDs for all 18 NN ensembles. The top-performing subset for ERA5-SSF clearly outperforms the other five subsets. For E3SM72 and E3SM×8, the top two best-performing subsets perform very similarly, but they stand out from the rest of the subsets. The ranks of the top three best-performing subsets suggested by these distributions are consistent with those determined by the NNs with the baseline architecture, shown in the legend in Figure 5. While it is certainly possible that the randomness in the training process (e.g., random initialization of NN parameters, order of data points fed to the NN, and so on) can change the top six best-performing subsets, it is unlikely that any other 8-MF subset will outperform the top-performers that we have examined here (i.e., the top-performer for ERA5-SSF and the top two best-performing subsets for E3SM72 and E3SM×8).

Table 1 shows the top six best-performing 8-MF subsets for the three data sets. θ_{850} and RH_{850} are present in all 18 8-MF subsets, closely followed by θ_{1000}, θ_{700}, and RH_{1000}, which make 15 to 16 appearances. RH_{700} is rarely useful and U_{700} is not used by any of these 18 8-MF subsets. Q_h and Q_e seem to be interchangeable. The best-performing 8-MF subsets for E3SM72 and E3SM×8 are similar. Actually, these two data sets share

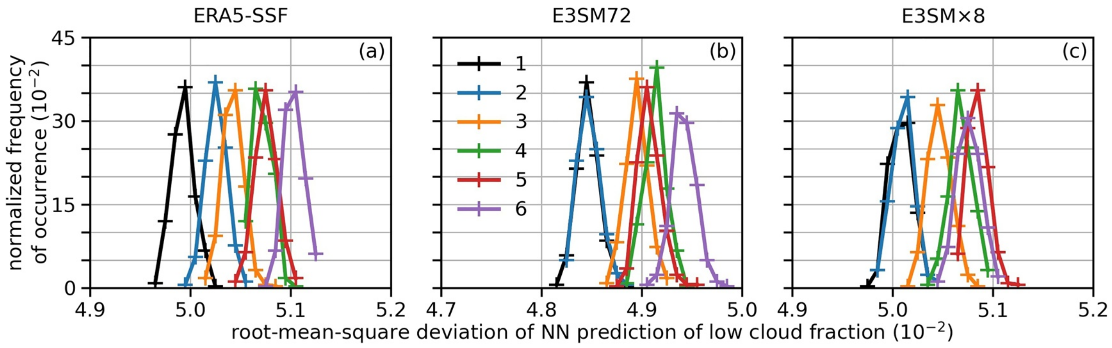

Figure 5. Performance of neural network (NN) ensembles generated with 341 perturbed NN architectures and trained to map the top six best-performing 8-MF subsets to the low cloud fractions. Numbers in the legend indicate the ranks of the 8-MF subsets determined by the performance of the NNs with the baseline architecture.

Table 1
Initial 14 Meteorological Factors (MFs) and Best-Performing 8-MF Subsets

Data set	ERA5-SSF						E3SM72						E3SM×8						Number of Occurrences
Rank	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	
θ_{1000}	●		○		○	○	○	●	○	○	○	○	●	○	○	○	○	○	16
θ_{925}		○		○	○					○	○	○				○		○	8
θ_{850}	●	○	○	○	○	○	○	●	○	○	○	○	●	○	○	○	○	○	18
θ_{700}	●	○	○	○			○	●	○	○	○		●	○	○	○	○	○	15
RH_{1000}		○		○	○		○	●	○	○	○	○	●	○	○	○	○	○	15
RH_{925}	●		○			○	○	●	○			○	●	○	○		○		11
RH_{850}	●	○	○	○	○	○	○	●	○	○	○	○	●	○	○	○	○	○	18
RH_{700}																	○		1
U_{1000}						○	○	●		○	○	○	●	○		○	○	○	11
U_{700}																			0
ω_{500}	●	○	○	○	○	○													6
PWV	●	○	○	○	○	○													6
Q_{h}			○	○		○	○		○		○	○		○	○	○			10
Q_{e}	●	○			○			●	○	○			●		○		○	○	10

Note. For each column, circles indicate the meteroloigical factors (MFs) used for this 8-MF subset. The 8-MF subset with filled circles are selected for further investigation.

the same top two best-performing subsets, which share seven MFs, but one contains Q_{h} and the other one contains Q_{e}. ERA5-SSF prefers different MF subsets from E3SM72 and E3SM×8. All top six best-performing subsets for ERA5-SSF contain ω_{500} and PWV, while only one contains U_{1000}. In contrast, none of the top six best-performing subsets for E3SM72 and E3SM×8 contains ω_{500} or PWV, but most of them contain U_{1000}. It is possible that different best-performing 8-MF subsets between E3SM-based data sets and ERA5-SSF indicate different physics between E3SM and the real world. However, keep in mind that while the LCFs and MFs from E3SM72 and E3SM×8 are generated by self-consistent systems, the SSF LCFs and ERA5 MFs are imperfect representations of the real world LCFs and MFs. Besides, the meteorological data used for retrieving SSF LCFs are not from ERA5 (See the descriptions in Minnis et al., 2020 or Trepte et al., 2019 for sources of data). So, the SSF LCFs may be inconsistent with both the real world meteorology and the ERA5 MFs. It is possible that ω_{500} and PWV better characterize the large-scale features that are shared among the real world, ERA5, and the data used for retrieving SSF LCFs, compared with MFs that characterize the details in the boundary layer.

We proceed with the following 8-MF subsets: θ_{1000}, θ_{850}, θ_{700}, RH_{925}, RH_{850}, Q_{e}, ω_{500}, and PWV for ERA5-SSF; θ_{1000}, θ_{850}, θ_{700}, RH_{1000}, RH_{925}, RH_{850}, Q_{e}, and U_{1000} for E3SM72 and E3SM×8. The latter 8-MF subset shares six MFs with the former, but is complemented by RH_{1000} and U_{1000} instead of ω_{500} and PWV. Figure 6 demonstrates the performance of the NN ensembles based on the selected 8-MF subsets. The first row displays the spatial patterns of the mean deviation between the true LCF and NN ensemble mean. Positive/negative deviation indicates that the NN ensemble mean underestimates/overestimates the truth. The mean deviations are mostly small (between ±0.03 at over 98.6% locations), suggesting satisfying overall performance of the NN ensembles. Those with large magnitudes are only occasionally found near the west coast of both the North and South Americas and to the south of the Hawaiian islands. It is possible that our exclusion of grids near landmasses is effective but does not completely eliminate the challenging conditions, where the topography near or within a grid may provide additional forcings—physical or numerical—that affect the LCF on the monthly timescale. We plot the ratio between the squared mean deviation and the mean squared deviation at each location (second row in Figure 6) and find that this ratio is usually small, suggesting that the NNs do not consistently produce deviations with large magnitude and a certain sign at a given location. It is much more difficult to examine whether NNs are consistently biased in some region in the 8-dimensional MF space. We check the correlation between the deviation and each MF in the selected 8-MF subset

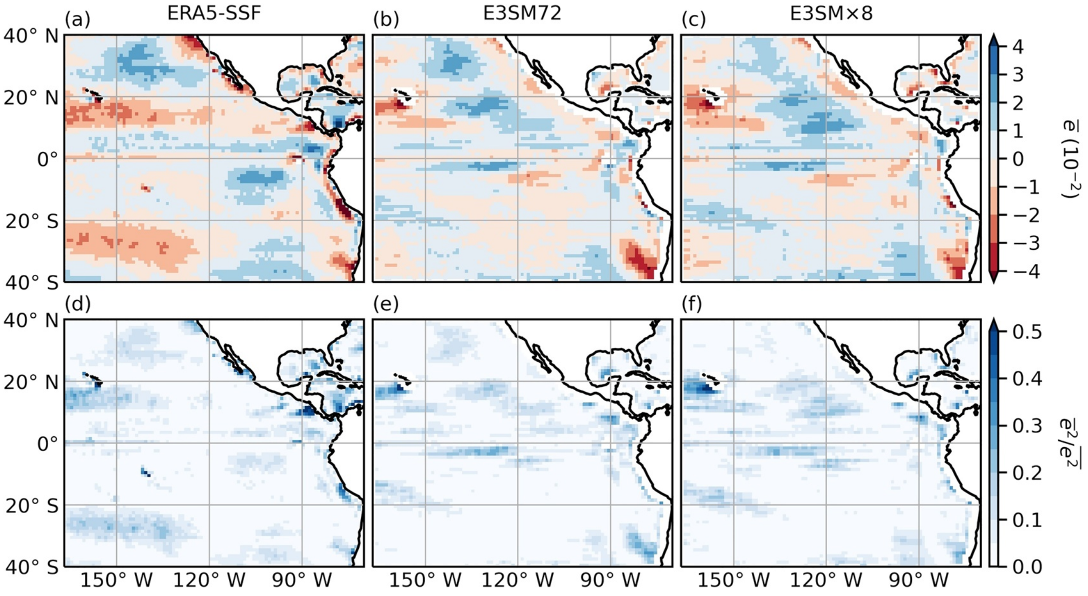

Figure 6. Performance of neural network (NN) ensembles trained on the selected 8-MF subsets. First row shows the mean of the deviation ($\bar{e}$, where the deviation, e, is defined as the true low cloud fractions minus the NN ensemble mean); second row shows the ratio between the squared mean deviation ($\bar{e}^2$) and the mean squared deviation ($\overline{e^2}$).

and find them to be uncorrelated. The correlation coefficient between the NN ensemble mean and the truth is greater than 0.966 for all data sets (plots omitted).

3.2. Classification of MFs

Next, we use the NN ensemble standard deviation to classify MFs into shared and uniquely occupied MF subspaces. Figure 7a shows the joint distribution of NN ensemble standard deviation and NN ensemble

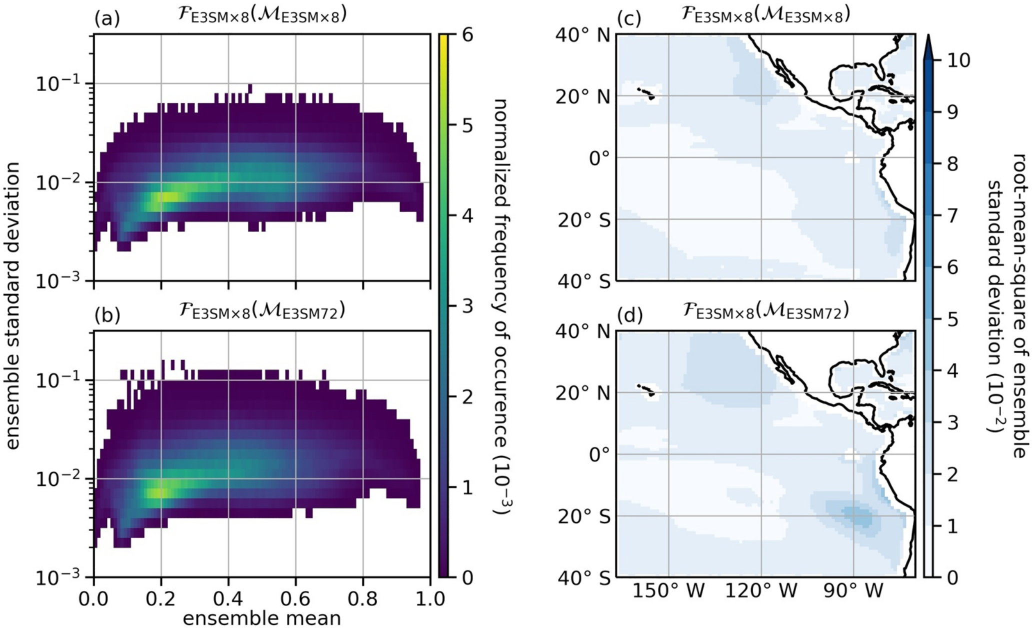

Figure 7. Joint distributions of neural network (NN) ensemble standard deviation and NN ensemble mean for (a) $\mathcal{F}_{\mathrm{E3SM\times8}}(\mathcal{M}_{\mathrm{E3SM\times8}})$ and (b) $\mathcal{F}_{\mathrm{E3SM\times8}}(\mathcal{M}_{\mathrm{E3SM72}})$; Spatial patterns of the root-mean-square of NN ensemble standard deviations for (c) $\mathcal{F}_{\mathrm{E3SM\times8}}(\mathcal{M}_{\mathrm{E3SM\times8}})$ and (d) $\mathcal{F}_{\mathrm{E3SM\times8}}(\mathcal{M}_{\mathrm{E3SM72}})$.

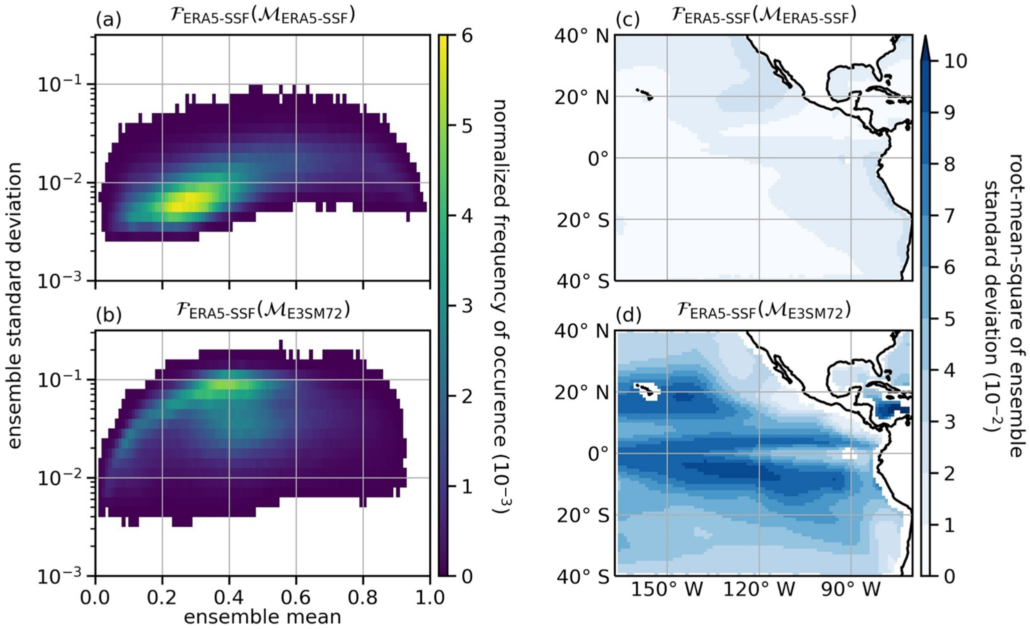

Figure 8. Same as Figure 7 but for E3SM72 and ERA5-SSF.

mean for $\mathcal{F}_{\text{E3SM}\times 8}(\mathcal{M}_{\text{E3SM}\times 8})$. The NN ensemble standard deviation is smallest for some data points with NN ensemble mean around 0.1, increases with the NN ensemble mean, and remains steady for NN ensemble mean between about 0.3 to 0.7. The largest standard deviation is smaller than 0.1. Recall that if we were to draw 341 random numbers independently from a uniform distribution between 0 and 1 and use them as the predicted LCFs, then the mean prediction would be close to 0.5 and the standard deviation close to 0.29. So, the NN ensemble members agree with one another significantly better than the randomly drawn numbers. Over the region of interest, the RMS of the NN ensemble standard deviation is smaller than 0.03 except for a few locations near the west coasts of South America (Figure 7c). When $\mathcal{F}_{\text{E3SM}\times 8}$ is applied to $\mathcal{M}_{\text{E3SM72}}$, the NN ensemble standard deviation is more dispersed, most evidently for NN ensemble mean up to 0.8 (Figure 7b). The RMS of the NN ensemble standard deviation becomes larger at most locations (Figure 7d), most evidently for two regions: one to the north and east of the Californian stratocumulus deck in E3SM72 and the other one centered around 20° S, 90° W to the south of the Peruvian stratocumulus deck in E3SM72.

For the data used in Figure 7a, we have validated that the NN ensemble means are close to the truth. However, the true values are unknown for the NN ensemble means shown in Figure 7b. We ignore the dependency of the NN ensemble standard deviation on the NN ensemble mean and compare the distributions of the NN ensemble standard deviation for $\mathcal{F}_{\text{E3SM}\times 8}(\mathcal{M}_{\text{E3SM}\times 8})$ and $\mathcal{F}_{\text{E3SM}\times 8}(\mathcal{M}_{\text{E3SM72}})$. Altogether, 79.9% of $\mathcal{M}_{\text{E3SM72}}$ produce $\mathcal{F}_{\text{E3SM}\times 8}(\mathcal{M}_{\text{E3SM72}})$ ensemble standard deviations smaller than 0.0188, which is the 90th percentile of the $\mathcal{F}_{\text{E3SM}\times 8}(\mathcal{M}_{\text{E3SM}\times 8})$ ensemble standard deviations. We classify these $\mathcal{M}_{\text{E3SM72}}$ into an MF subspace that is shared between E3SM×8 and E3SM72, and the rest of $\mathcal{M}_{\text{E3SM72}}$ into the MF subspace that is uniquely occupied by E3SM72. In Figure 9a, we find smaller fractions of $\mathcal{M}_{\text{E3SM72},\cap\text{E3SM}\times 8}$ in the regions that see larger increases in the RMS of NN ensemble standard deviation from Figure 7c to Figure 7d. Similarly, we use the $\mathcal{F}_{\text{E3SM72}}$ ensemble to classify $\mathcal{M}_{\text{E3SM}\times 8}$. The spatial pattern of the fraction of $\mathcal{M}_{\text{E3SM}\times 8,\cap\text{E3SM72}}$ (Figure 9b) is similar to that for $\mathcal{M}_{\text{E3SM72},\cap\text{E3SM}\times 8}$.

Figure 8 shows the same results but for ERA5-SSF and E3SM72. Although the patterns in Figures 8a and 8c are comparable with those in Figures 7a and 7c, NN ensemble standard deviations for $\mathcal{F}_{\text{ERA5-SSF}}(\mathcal{M}_{\text{E3SM72}})$ are much greater than those for $\mathcal{F}_{\text{ERA5-SSF}}(\mathcal{M}_{\text{ERA5-SSF}})$. In Figure 8b, the highest frequency of occurrence is found near an NN ensemble standard deviation close to 0.1 and an NN ensemble mean close to 0.4, different from those in Figures 7a, 7b and 8a, indicating generally large NN ensemble standard deviations. While the largest RMS of the $\mathcal{F}_{\text{E3SM}\times 8}(\mathcal{M}_{\text{E3SM72}})$ ensemble standard deviation is smaller than 0.05 (Figure 7d), that for $\mathcal{F}_{\text{ERA5-SSF}}(\mathcal{M}_{\text{E3SM72}})$ is greater than 0.1. Comparing the spatial patterns in Figures 8c and 8d, the greatest

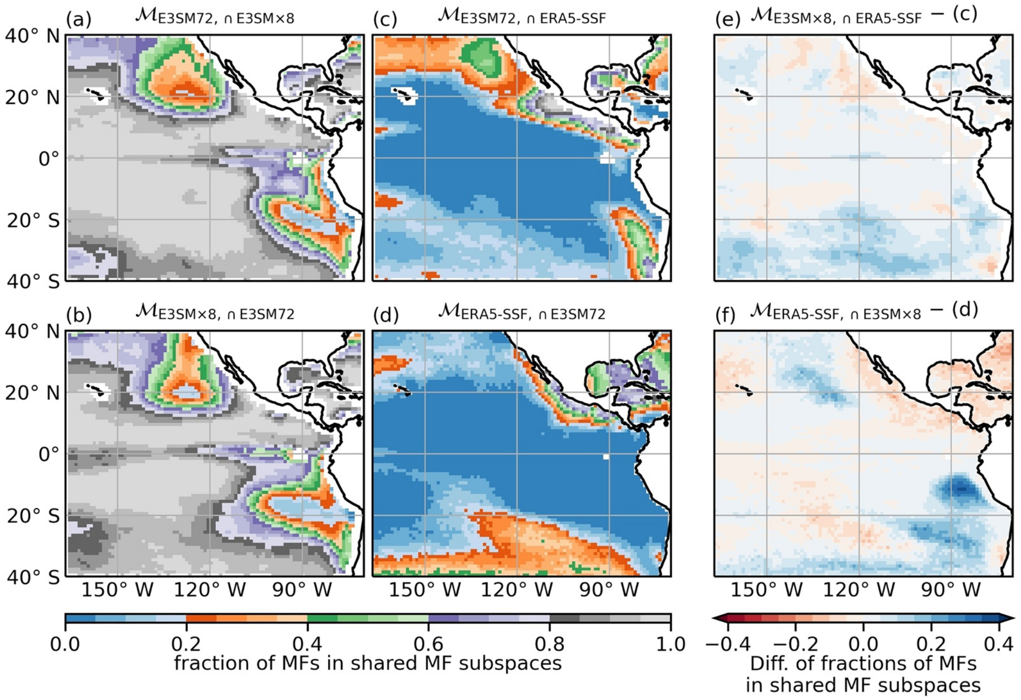

Figure 9. (a–d) Fractions of meteorological factors (MFs) in shared MF subspaces; (e) difference between fractions of $\mathcal{M}_{\mathrm{E3SM\times8},\cap\mathrm{ERA5\text{-}SSF}}$ and $\mathcal{M}_{\mathrm{E3SM72},\cap\mathrm{ERA5\text{-}SSF}}$; (f) difference between fractions of $\mathcal{M}_{\mathrm{ERA5\text{-}SSF},\cap\mathrm{E3SM\times8}}$ and $\mathcal{M}_{\mathrm{ERA5\text{-}SSF},\cap\mathrm{E3SM72}}$.

increases in the RMS of NN ensemble standard deviations from $\mathcal{F}_{\mathrm{ERA5\text{-}SSF}}(\mathcal{M}_{\mathrm{ERA5\text{-}SSF}})$ to $\mathcal{F}_{\mathrm{ERA5\text{-}SSF}}(\mathcal{M}_{\mathrm{E3SM72}})$ occur over the tropical oceans far away from continents, different from the case for E3SM×8 and E3SM72. Only 13.8% of the $\mathcal{F}_{\mathrm{ERA5\text{-}SSF}}(\mathcal{M}_{\mathrm{E3SM72}})$ ensemble standard deviations are smaller than 0.0192, the 90^{th} percentile of the $\mathcal{F}_{\mathrm{ERA5\text{-}SSF}}(\mathcal{M}_{\mathrm{ERA5\text{-}SSF}})$ ensemble standard deviations. The very limited overlap between the MF subspace occupied by $\mathcal{M}_{\mathrm{E3SM72}}$ and $\mathcal{M}_{\mathrm{ERA5\text{-}SSF}}$ is also reflected in the spatial pattern of the fraction of $\mathcal{M}_{\mathrm{E3SM72},\cap\mathrm{ERA5\text{-}SSF}}$ (Figure 9c) and that of $\mathcal{M}_{\mathrm{ERA5\text{-}SSF},\cap\mathrm{E3SM72}}$ (Figure 9d).

The results for ERA5-SSF and E3SM×8 are comparable to those for ERA5-SSF and E3SM72. We highlight the differences between the fractions of $\mathcal{M}_{\mathrm{E3SM\times8},\cap\mathrm{ERA5\text{-}SSF}}$ and $\mathcal{M}_{\mathrm{E3SM72},\cap\mathrm{ERA5\text{-}SSF}}$ (Figure 9e) and the differences between the fractions of $\mathcal{M}_{\mathrm{ERA5\text{-}SSF},\cap\mathrm{E3SM\times8}}$ and $\mathcal{M}_{\mathrm{ERA5\text{-}SSF},\cap\mathrm{E3SM72}}$ (Figure 9f). It is clear that the overlap between $\mathcal{M}_{\mathrm{ERA5\text{-}SSF}}$ and $\mathcal{M}_{\mathrm{E3SM\times8}}$ is greater than that between $\mathcal{M}_{\mathrm{ERA5\text{-}SSF}}$ and $\mathcal{M}_{\mathrm{E3SM72}}$ for both the North and the Southeast Pacific. The largest increase occurs in the region centered around 13° S, 87° W, where the fraction of $\mathcal{M}_{\mathrm{ERA5\text{-}SSF},\cap\mathrm{E3SM\times8}}$ can be more than 30% larger than the fraction of $\mathcal{M}_{\mathrm{ERA5\text{-}SSF},\cap\mathrm{E3SM72}}$. The differences in patterns in Figures 9e and 9f are expected since the distributions of MFs from each pair of data sets (e.g., $\mathcal{M}_{\mathrm{ERA5\text{-}SSF}}$ and $\mathcal{M}_{\mathrm{E3SM72}}$) are different in their shared MF subspace.

3.3. Decomposition of LCF Differences

In this subsection, we decompose the differences in LCF between the three data sets into those associated with different MFs and with different LCF-MF relationships.

We calculate the mean patterns of $f_{\mathrm{E3SM\times8},\cap\mathrm{E3SM72}}$ and $f_{\mathrm{E3SM72},\cap\mathrm{E3SM\times8}}$ and the difference between them, and then repeat the calculation for $f_{\mathrm{E3SM\times8},\cap\overline{\mathrm{E3SM72}}}$ and $f_{\mathrm{E3SM72},\cap\overline{\mathrm{E3SM\times8}}}$. Qualitatively, the pattern of the difference between $f_{\mathrm{E3SM\times8},\cap\mathrm{E3SM72}}$ and $f_{\mathrm{E3SM72},\cap\mathrm{E3SM\times8}}$ (Figure 10a) correlates well with the difference between $f_{\mathrm{E3SM\times8}}$ and f_{E3SM72} (Figure 2f), with E3SM×8 having greater LCF to the north of 20° N and near the Peruvian stratocumulus deck. However, the magnitude of the difference between $f_{\mathrm{E3SM\times8},\cap\mathrm{E3SM72}}$ and $f_{\mathrm{E3SM72},\cap\mathrm{E3SM\times8}}$ is smaller than that in Figure 2f. Also, there is a clear offset between the areas with most positive differences in Figure 10a and in Figure 2f, with the former residing to the north and east of 20° S, 90° W and the latter centered around that point. The pattern of the difference between $f_{\mathrm{E3SM\times8},\cap\overline{\mathrm{E3SM72}}}$ and $f_{\mathrm{E3SM72},\cap\overline{\mathrm{E3SM\times8}}}$ also

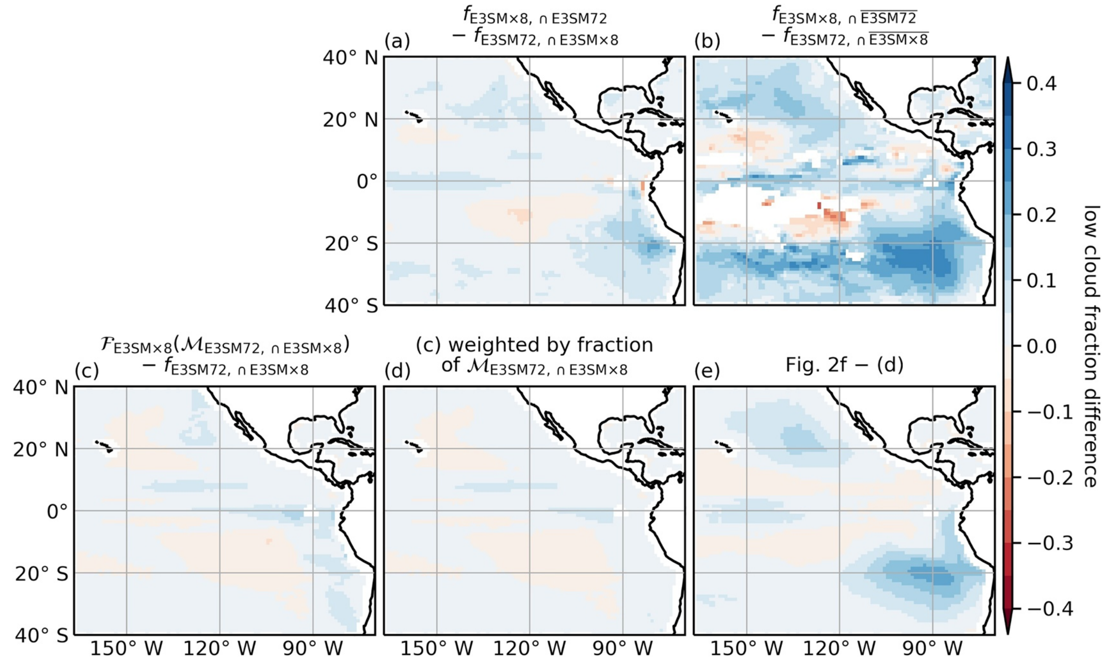

Figure 10. Spatial patterns of difference between mean $f_{\mathrm{E3SM\times8}}$ and f_{E3SM72} associated with meteorological factors (MFs) in (a) shared and (b) uniquely occupied MF subspaces; spatial patterns of difference between $\mathcal{F}_{\mathrm{E3SM\times8}}$ and f_{E3SM72} for $\mathcal{M}_{\mathrm{E3SM72},\cap\mathrm{E3SM\times8}}$ (c) before and (d) after being weighted by the fraction of $\mathcal{M}_{\mathrm{E3SM72},\cap\mathrm{E3SM\times8}}$; (e) spatial pattern of remaining difference between $f_{\mathrm{E3SM\times8}}$ and f_{E3SM72}.

shows positive differences in the North and Southeast Pacific (Figure 10b) with magnitudes greater than those in Figure 10a and the greatest values to the south of 20° S.

Note that the sum of the differences shown in Figures 10a and 10b is not equal to the differences in Figure 2f. Also, although $f_{\mathrm{E3SM\times8},\cap\mathrm{E3SM72}}$ and $f_{\mathrm{E3SM72},\cap\mathrm{E3SM\times8}}$ are based on MFs in the shared MF subspace, one is associated with $\mathcal{M}_{\mathrm{E3SM\times8},\cap\mathrm{E3SM72}}$ and the other one with $\mathcal{M}_{\mathrm{E3SM72},\cap\mathrm{E3SM\times8}}$, the joint distribution of which within the shared MF subspace could be different from that of $\mathcal{M}_{\mathrm{E3SM\times8},\cap\mathrm{E3SM72}}$. To remove these two factors from the comparison, we first take an intermediate step to examine the difference between the mean patterns of $\mathcal{F}_{\mathrm{E3SM\times8}}(\mathcal{M}_{\mathrm{E3SM72},\cap\mathrm{E3SM\times8}})$ ensemble mean and $f_{\mathrm{E3SM72},\cap\mathrm{E3SM\times8}}$. Shown in Figure 10c, these relatively small LCF differences with values between –0.05 and 0.10 at most locations are entirely due to the difference in the LCF-MF relationship between E3SM×8 and E3SM72 for $\mathcal{M}_{\mathrm{E3SM72},\cap\mathrm{E3SM\times8}}$. The magnitude of the differences in LCF from this source gets slightly smaller after being weighted by the fraction of $\mathcal{M}_{\mathrm{E3SM72},\cap\mathrm{E3SM\times8}}$ at each location (Figure 10d). The remaining differences between $f_{\mathrm{E3SM\times8}}$ and f_{E3SM72}, shown in Figure 10e, involve a shift from E3SM72 MFs to E3SM×8 MFs. We conclude that most of the differences between $f_{\mathrm{E3SM\times8}}$ and f_{E3SM72} are associated with different MFs between the two data sets.

We repeat the comparison for E3SM72 and ERA5-SSF (Figure 11). The magnitude of differences between $f_{\mathrm{E3SM72},\cap\mathrm{ERA5\text{-}SSF}}$ and $f_{\mathrm{ERA5\text{-}SSF},\cap\mathrm{E3SM72}}$ (Figures 11a) is much greater than that for $f_{\mathrm{E3SM\times8},\cap\mathrm{E3SM72}}$ and $f_{\mathrm{E3SM72},\cap\mathrm{E3SM\times8}}$ (Figure 10a) but smaller than that between $f_{\mathrm{E3SM72},\cap\overline{\mathrm{ERA5\text{-}SSF}}}$ and $f_{\mathrm{ERA5\text{-}SSF},\cap\overline{\mathrm{E3SM72}}}$ (Figures 11b). The differences between $f_{\mathrm{E3SM72},\cap\mathrm{ERA5\text{-}SSF}}$ and $f_{\mathrm{ERA5\text{-}SSF},\cap\mathrm{E3SM72}}$ at some locations may not be representative because of low fractions of MFs in the shared MF subspace. In the North Pacific, only a few locations very close to the North American coast see relatively high fractions of both $\mathcal{M}_{\mathrm{E3SM72},\cap\mathrm{ERA5\text{-}SSF}}$ and $\mathcal{M}_{\mathrm{ERA5\text{-}SSF},\cap\mathrm{E3SM72}}$. The fraction of $\mathcal{M}_{\mathrm{ERA5\text{-}SSF},\cap\mathrm{ERA5\text{-}SSF}}$ quickly drops to less than 20% (38 data points), 10% (19 data points), and even 5% (10 data points) as one moves toward the south and the west. Similarly, the relatively large negative differences in the South Pacific are based on a few data points from ERA5-SSF. Still, the mean differences between $f_{\mathrm{E3SM72},\cap\mathrm{ERA5\text{-}SSF}}$ and $\mathcal{F}_{\mathrm{ERA5\text{-}SSF}}(\mathcal{M}_{\mathrm{E3SM72},\cap\mathrm{ERA5\text{-}SSF}})$ ensemble mean can easily reach below –0.15 around the Californian stratocumulus deck and –0.30 in the Southeast Pacific. It is not surprising that the magnitude of the differences between $f_{\mathrm{E3SM72},\cap\mathrm{ERA5\text{-}SSF}}$ and $\mathcal{F}_{\mathrm{ERA5\text{-}SSF}}(\mathcal{M}_{\mathrm{E3SM72},\cap\mathrm{ERA5\text{-}SSF}})$ in Figures 11d is small, since the results are weighted by the small fraction of $\mathcal{M}_{\mathrm{E3SM72},\cap\mathrm{ERA5\text{-}SSF}}$. Consequently,

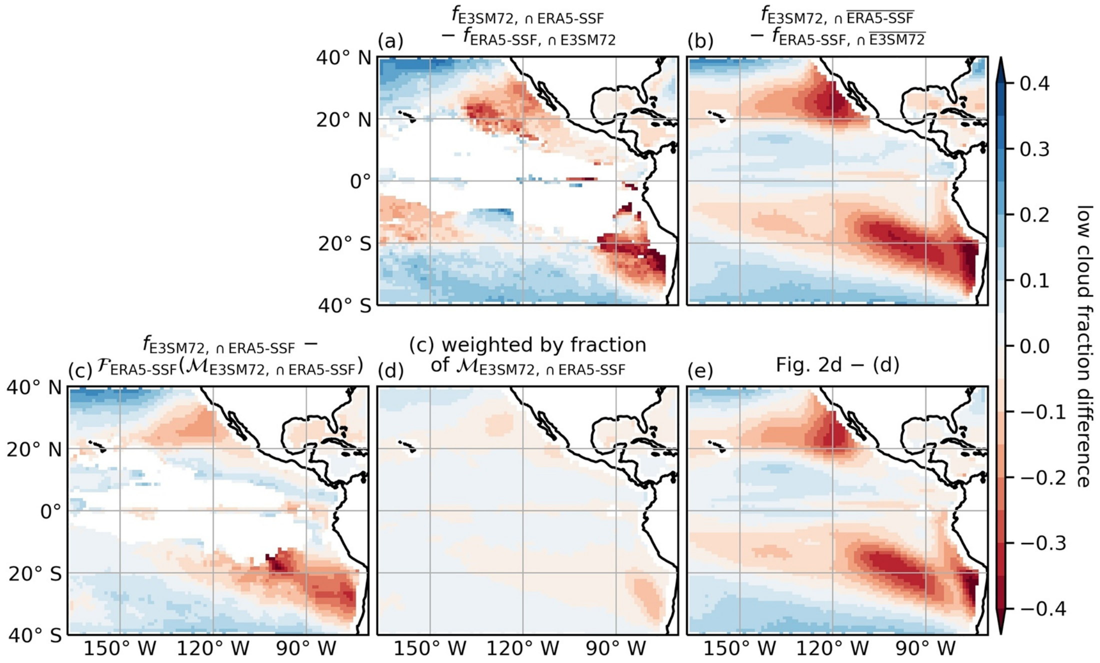

Figure 11. Same as Figure 10 but for comparison between ERA5-SSF and E3SM72.

the remaining differences that are associated with a shift from ERA5 MFs to E3SM72 MFs (Figures 11e) dominate the total differences between f_{E3SM72} and $f_{\mathrm{ERA5\text{-}SSF}}$ (Figure 2d).

The results for E3SM×8 and ERA5-SSF are again comparable to those for E3SM72 and ERA5-SSF and omitted here.

4. Discussion

4.1. Implications of LCF Decomposition Results

Regarding the model intercomparison between E3SM72 and E3SM×8, we discover that the environments for marine low clouds in E3SM72 and E3SM×8 (characterized by the selected 8 MFs) differ around both the Californian and Peruvian stratocumulus decks, and most of the difference between $f_{\mathrm{E3SM\times8}}$ and f_{E3SM72} is associated with a shift in MFs, detected by NN ensemble standard deviations. Our results suggest that there are twofold impacts of using 8-times vertical resolution via FIVE in the lower troposphere for certain processes (i.e., three physical parameterizations and the vertical advection by the resolved flow). On the one hand, E3SM×8 directly produces greater LCFs than E3SM72 given the same MFs in the shared MF subspace. On the other hand, E3SM×8 shifts the low cloud environments toward meteorological regimes, characterized by the MFs, that support higher LCF. The latter indirect impact has to be caused by the processes computed on the high vertical resolution feeding back to the standard resolution aspect of E3SM-FIVE. In our results, this indirect impact dominates the difference between $f_{\mathrm{E3SM\times8}}$ and f_{E3SM72}. However, the shift in MFs is confined to limited regions, suggesting a weak coupling between the stratocumulus and cumulus regimes and the boundary layer over other oceanic areas in our region of interest in E3SM's atmospheric component.

Regarding the evaluation of E3SM72 and E3SM×8 against ERA5-SSF, we find large discrepancies in the LCF-MF relationship between the E3SM-based data sets and ERA5-SSF. In this sense, our results agree with previous studies that physical parameterizations are responsible for errors in E3SM-produced LCFs. Recall that the primary reason to compute the vertical advection by the resolved flow at the high vertical resolution is to balance the parameterized entrainment due to turbulence; we consider this component as part of the physical parameterizations that are different between E3SM×8 and E3SM72. However, we show that the differences in the MFs are still large enough to dominate the differences between E3SM-based LCFs and SSF

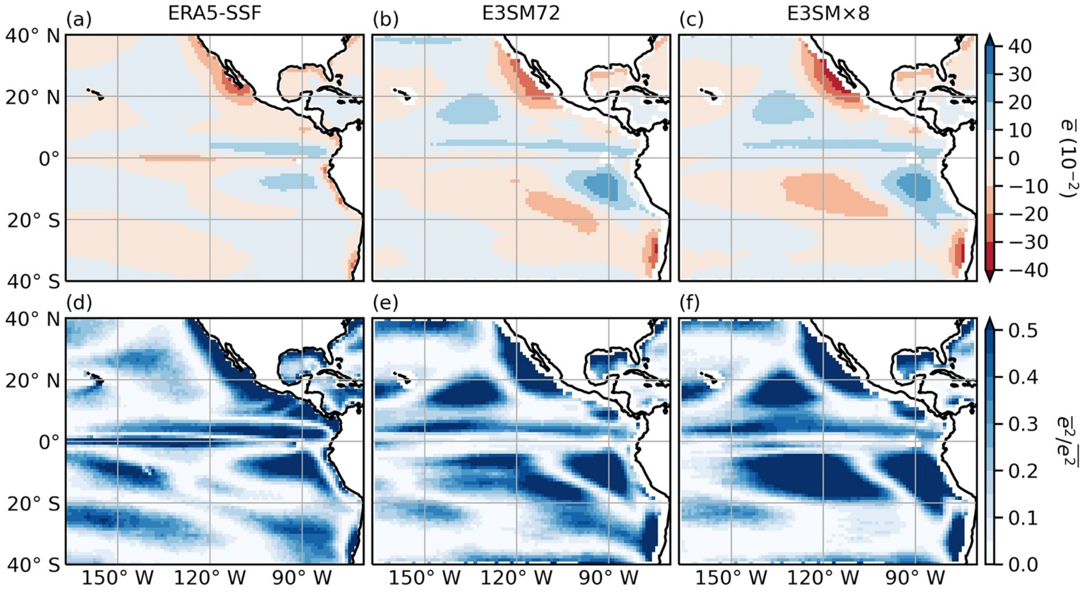

Figure 12. Same as Figure 6 but for a single neural network with the baseline architecture and trained on two meteorological factors: θ_{700} and θ_{1000}.

LCFs. What we are unable to conclude from the current study is whether (and how much) a set of "perfect" parameterizations for the processes parameterized by CLUBB, MG2, and RRTMG alone is sufficient to shift the MFs in E3SM toward more realistic meteorological regimes.

In the following subsections, we discuss a few issues that relate to our methods and may affect the robustness of our results.

4.2. Considerations Behind the MF Pool

We have taken a brute-force approach to pick the best-performing features for each data set. Once we make the decision to settle on 8-MF subsets, this feature selection method guarantees delivery of a global optimum MF subset given the MF candidates, albeit at a relatively high computational cost. While training NNs on 16,383 MF subsets and a couple of hundred ensemble members for selected MF subsets is manageable, the task quickly becomes unattainable as the MF pool expands. Nevertheless, we test 8-MF subsets for a few slightly perturbed MF pools. We find that replacing ω_{500} with ω_{700} decreases the performance of $\mathcal{F}_{\text{ERA5-SSF}}$; like ω_{500}, ω_{700} does not make it into the best-performing subsets for E3SM72 and E3SM×8.

One limitation of our NN approach is that it captures the correlation between MFs and LCF but does not reveal the causation. So, highly correlated MFs will be interchangeable. We do not test surface temperature advection due to its correlation with surface sensible heat flux, nor SST due to its correlation with θ_{1000}. Weaver and Ramanathan (1997) suggested that the monthly ω_{500} variance is a good predictor of the cloud radiative forcing (CRF) over the Northern Hemisphere extratropical oceans. Despite the correlation between CRF and LCF, we do not find any advantage in using ω_{500} variance as a feature, probably because other MFs we examine contain sufficient information to predict the LCF over extratropical oceans.

A related issue is the need to avoid MFs that are too closely related to, or even equivalent to the LCF. We argue that it is inevitable to include MFs related to the humidity in the lower troposphere, simply because moisture is an essential component of cloud formation. We examine the joint distributions between the LCF and a few humidity related MFs (i.e., RH_{1000}, RH_{925}, RH_{850}, and PWV) in the form of two-dimensional histograms. The patterns can be very different for the same MF across three data sets. While these different joint distributions themselves can be used to diagnose model defects (Bennhold & Sherwood, 2008), for our work we believe that the success of our NNs does not rest on trivially simple relationships between humidity-related MFs and the LCF. Another MF we examine is the solar insolation at the top-of-atmosphere (SOLIN).

$\mathcal{F}_{\text{ERA5-SSF}}$ performs better when SOLIN is included as a feature. However, we drop SOLIN so that we are not mapping a strong indicator of the seasonality to the LCF, which itself experiences seasonal changes.

4.3. Comparison With LTS, EIS, ECTEI, and ELF

In this work, we have selected the best-performing MF subsets from basic MFs. It is natural to ask (a) how the 8-MF subsets we select compare with compound MFs that have been shown to correlate well with the LCFs, for example, LTS, EIS, ECTEI, and ELF and (b) whether it is beneficial to include these compound MFs in the initial MF pool. The first question can be partially answered with Figures 5 and 8 in Park and Shin (2019), where the correlation coefficients reported suggest that the linear statistical models that map each of these compound MFs to the LCF do not outperform the NN ensembles. However, there are differences between their and our data and methods (e.g., spatial and temporal scales). The second question is still a valid one, since it is possible that an NN trained on one of these four compound MFs plus only a few extra basic MFs might achieve the same performance as an NN trained on more basic MFs, thus reducing the dimensionality of the problem.

To answer these two questions, we trained additional NNs to map three new sets of MFs to the LCFs, for each data set, and for each compound MF. MF Set 1 is just the compound MF itself. MF Set 2 is composed of the MFs that are used to construct the compound MF. We define LTS as $\theta_{700} - \theta_{1000}$, so the MFs used to construct LTS are simply θ_{1000} and θ_{700}. EIS is estimated from θ_{1000}, θ_{700}, and RH_{1000}, slightly different from Wood and Bretherton (2006), where the authors used surface temperature, surface pressure, and a constant RH of 80%. ECTEI is calculated from θ_{1000}, θ_{700}, RH_{1000}, and RH_{700}, following Kawai et al. (2017), and only involves one more MF than EIS. ELF is also estimated from θ_{1000}, θ_{700}, and RH_{1000}, the same MFs that are used to calculate EIS, based on the first line in Equation 4 in Park and Shin (2019) and similar simplifications adopted in Wood and Bretherton (2006). Including θ and RH from more levels should improve the estimation of the inversion base height and the surface-based mixing layer top height, the two key components for ELF. Thus, we define an MF subset comprising θ and RH at 1,000, 925, 850, and 700 mb as the MF Set 2 for ELF to differentiate from the MF Set 2 for EIS. MF Set 3 is composed of the compound MF plus the MFs in the selected 8-MF subset but not included in MF Set 2. For example, MF Set 3 for ERA5-SSF for ELF contains ELF (the compound MF itself) and Q_e, ω_{500}, and PWV (because other MFs in the selected 8-MF subset for ERA5-SSF are present in MF Set 2 for ERA5-SSF for ELF). Only the NNs with baseline architectures are trained on these new MF sets.

The results of these tests show that none of these alternative MF sets outperforms the selected 8-MF subsets. The NN trained on MF Set 2 for ELF performs well, but this MF set is just one of the 8-MF subsets that we examined before and we already know that it is not the top-performer. The performance of the NNs trained on MF Set 3 for LTS, EIS, or ECTEI are close to the selected 8-MF subsets. However, these NNs are trained on six or seven features, only one or two fewer than the 8-MF subsets. Also, these MF sets do not outperform the best-performing 6-MF or 7-MF subsets shown in Figure 3. We conclude that there is no obvious advantage in including these compound MFs in our initial MF pool.

4.4. Other Issues

A key component of our work is to classify MFs into shared and uniquely occupied MF subspaces. This task falls into the broader category of problems known as outlier detection or anomaly detection for which many techniques have been developed. See a few examples summarized in Géron, 2019, Chapter 9. The task becomes more complicated in high-dimensional space (Zimek et al., 2012). We tackle this task by considering NN ensemble standard deviations. Although this method has not been reported in the scientific literature, the philosophy behind it is sound: if a new data point falls in the area where we are not able to sufficiently constrain the NN ensemble members, it probably does not belong to the data used to train the NN ensemble. Regardless, one can imagine that some data points may be ambiguous enough to cause uncertainty in the classification results. While we use the toy problem (in Appendix A) to demonstrate the validity of this method, more rigorous studies are warranted to provide a more solid foundation.

We comment on two aspects of NN ensemble standard deviation. First, the NN ensemble standard deviation when the NN ensemble is applied to its training data represents the intrinsic variability of NN ensemble member performance. This variability exists because there is no guarantee that the optimizer (ADAM in our case) that searches for the optimal NN parameters can find the global optimum, and consequently, the performance of each NN ensemble member is a random variable. Second, although we observe in the toy problem that the NN ensemble standard deviation is larger when the NN ensemble is applied to inputs outside of the range of its training data, it is possible that for some of these inputs, the NN ensemble standard deviation can accidentally be small, no matter whether the NN ensemble means prediction is close to the truth. These inputs will be classified as sharing the range with the training data. Following this logic, the fraction of MFs in the shared MF subspace reported in this work may be overestimated, but this does not affect our conclusion that the differences in MFs have a significant contribution to the differences in LCF.

We perform sensitivity tests regarding a few elements in our method. First, there are many methods for generating NN ensembles (see Section 7.11 in Goodfellow et al., 2016) and we test one alternative method to generate NN ensembles by simply training the NNs with the baseline architecture on selected 8-MF subsets 341 times for each data set. The random initialization of parameters leads to one unique NN each time. The distributions of the performances of NN ensemble members generated this way are narrower than those shown in Figure 5 but this does not change our conclusions. Second, we test two alternative values of the threshold used to classify the MFs (85% and 95%) and get similar results to those shown thus far (based on a threshold of 90%). Third, increasing the number of features will improve the performance of the NNs but may result in reduced fractions of MFs in shared MF subspaces. We repeat the analysis with NN ensembles (containing 226 NN ensemble members identified following the method in Section 2.3) trained on all 14 MFs. Still, the results are similar to what we find with selected 8-MF subsets.

We are essentially interpreting $\mathcal{M}_{\mathrm{A}}$ that produce relatively small $\mathcal{F}_{\mathrm{B}}(\mathcal{M}_{\mathrm{A}})$ ensemble standard deviation, compared with $\mathcal{F}_{\mathrm{B}}(\mathcal{M}_{\mathrm{B}})$ ensemble standard deviations, as sharing an MF subspace with $\mathcal{M}_{\mathrm{B}}$, where A is one of ERA5-SSF, E3SM72, and E3SM×8 and B is one of the remaining two. A fundamental question is whether this interpretation is reasonable. For example, we show in Figure 9a the spatial distribution of the frequency of occurence of $\mathcal{M}_{\mathrm{E3SM72}}$ that produce $\mathcal{F}_{\mathrm{E3SM\times8}}(\mathcal{M}_{\mathrm{E3SM72}})$ smaller than the 90th percentile of $\mathcal{F}_{\mathrm{E3SM\times8}}(\mathcal{M}_{\mathrm{E3SM\times8}})$. While we interpret it as the spatial distribution of the fraction of $\mathcal{M}_{\mathrm{E3SM72,\cap E3SM\times8}}$, one can also interpret the same figure as the spatial distribution of the frequency of occurrence of $\mathcal{M}_{\mathrm{E3SM72}}$ that produce $\mathcal{F}_{\mathrm{E3SM\times8}}(\mathcal{M}_{\mathrm{E3SM72}})$ smaller than the 80th percentile of $\mathcal{F}_{\mathrm{E3SM\times8}}(\mathcal{M}_{\mathrm{E3SM72}})$, since the fraction of $\mathcal{M}_{\mathrm{E3SM72,\cap E3SM\times8}}$ is about 80%. We plot the spatial distribution of the frequency of occurrence of $\mathcal{M}_{\mathrm{E3SM72}}$ that produce $\mathcal{F}_{\mathrm{E3SM72}}(\mathcal{M}_{\mathrm{E3SM72}})$ smaller than the 80th percentile of $\mathcal{F}_{\mathrm{E3SM72}}(\mathcal{M}_{\mathrm{E3SM72}})$ and the frequency of occurrence of $\mathcal{M}_{\mathrm{E3SM\times8}}$ that produce $\mathcal{F}_{\mathrm{E3SM\times8}}(\mathcal{M}_{\mathrm{E3SM\times8}})$ smaller than the 80th percentile of $\mathcal{F}_{\mathrm{E3SM\times8}}(\mathcal{M}_{\mathrm{E3SM\times8}})$. These spatial distributions are significantly different from the patterns shown in Figure 9a, more so for the Southeast Pacific, confirming that we have been detecting anomalies, instead of identifying regions where the NN ensemble members naturally show larger standard deviations.

Last but not least, besides the different definitions of LCFs that we have discussed earlier, the definitions of MFs may not be the same between E3SM-based data sets and ERA5. For example, the model used to produce ERA5 runs with a 137-level vertical grid, different from E3SM's vertical grid, and different interpolation methods are used to interpolate variables from the two models' native vertical grids to selected pressure levels. For now, we simply assume that for a given variable, whatever is available from each data set is the best estimate of this variable for this data set.

5. Summary

Recent work has shown that FIVE improves the representation of shallow clouds in E3SM by computing selected physical parameterizations and the vertical advection by the resolved flow on high resolution vertical grids. In this study, we compare the results from an E3SM simulation on its default 72-layer vertical grid (E3SM72) with those from an E3SM-FIVE simulation using 8-times vertical resolution for the lower troposphere (E3SM×8) and evaluate these two simulations with ERA5 reanalysis and CERES SSF products (ERA5-SSF). Specifically, the goal is to understand the contributions of the MFs and the relationship between the LCF and the MFs to the improvement of LCFs from E3SM72 to E3SM×8, and the discrepancy between E3SM-produced LCFs and SSF LCFs.

The critical tool used is the NN ensemble. An NN ensemble is trained for each data set to not only capture the LCF-MF relationship but also to serve as proxy models to (a) predict LCFs for MFs from all data sets and (b) to classify MFs into those in shared and uniquely occupied MF subspaces. The idea is that the standard deviations among NN ensemble members when the NN ensemble trained on one data set is applied to the MFs from the same data set can be used as a reference to check the NN ensemble standard deviations when the NN ensemble is applied to MFs from another data set. MFs producing large NN ensemble standard deviations relative to the reference are more likely to fall outside the MF subspace that is occupied by MFs from the original data set. We apply this method to data over the oceanic region between 40° S, 40° N, 70° W, and 167° W.

Overall, E3SM72 and E3SM×8 have large fractions of MFs in shared MF subspace, but less so near the Californian and Peruvian stratocumulus decks. Both E3SM72 and E3SM×8 have small fractions of MFs in MF subspace shared with ERA5, but the fractions shared between E3SM×8 and ERA5 MFs are larger than those between E3SM72 and ERA5 MFs, especially in the Southeast Pacific. Using 8-times vertical resolution via FIVE in the lower troposphere for certain processes directly produces greater LCFs given the same (shared) MFs; it also shifts the low cloud environments toward meteorological regimes that support higher LCF. The latter impact dominates the difference in LCF between E3SM×8 and E3SM72. However, this impact is confined to limited regions, suggesting a weak coupling between the stratocumulus and cumulus regimes and the boundary layer over other oceanic areas in our region of interest in E3SM's atmospheric component. Both E3SM72 and E3SM×8 produce lower LCF than SSF given the same MFs in the very limited shared MF subspaces. However, it is the differences in the MFs between the E3SM-based data sets and ERA5-SSF that dominate the difference in LCF between them. Further improvement in E3SM-produced LCFs will require the improvement of physical parameterizations to narrow the gap between the LCF-MF relationships in the E3SM and ERA5 worlds. In addition, model development to bring E3SM-produced MFs closer to the observed MFs should be a key focus of effort.

Appendix A: Toy Problem Demonstrating the Behavior of NN Ensembles

We train an NN ensemble to fit a sinusoidal function $Y = \sin(X)$. The inputs for the training data set are randomly drawn between $-\pi$ and π. The NN ensemble consists 24 members, which all have the same architecture (two hidden layers with 10 hidden nodes in each layer) and are generated by randomly initializing weights and biases. Then the NN ensemble is used to make prediction for inputs from -3π to 3π. As shown

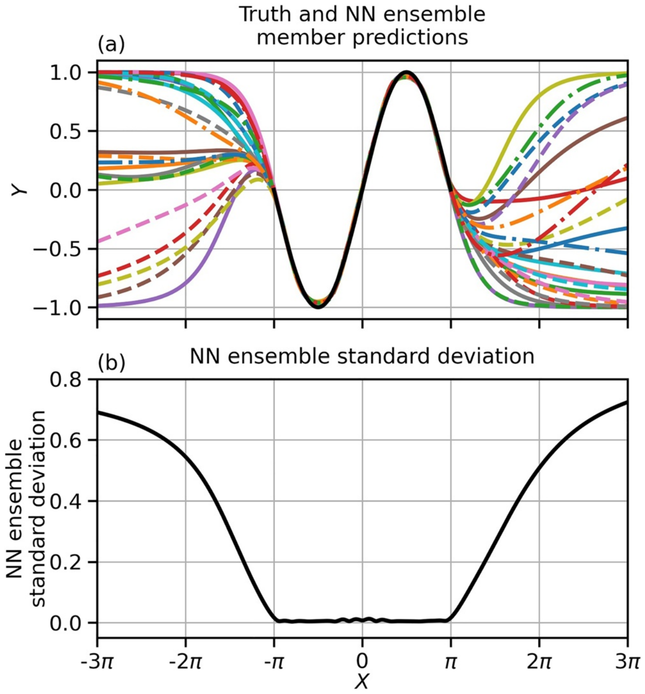

Figure A1. (a) Truth ($Y = \sin(X)$ between $-\pi$ and π, black solid line) and the predictions by 24 neural network (NN) ensemble members for inputs from -3π to 3π (other lines with various colors and line types). (b) Standard deviation among NN ensemble members.

in Figure A1a, predictions from all NN ensemble members agree reasonably well with the truth for inputs between $-\pi$ and π, where the NN ensemble is constrained by the training data, but diverge for inputs outside of this range. Consistent with this behavior, the standard deviation among NN ensemble members increases outside of $(-\pi, \pi)$ (Figure A1b).

Data Availability Statement

The data used for this research, including processed ERA5 MFs and SSF LCFs, are available at https://csl.noaa.gov/groups/csl9/datasets/data/cloud_phys/2021-Chen-etal/. The original ERA5 reanalysis data are available at the Copernicus Climate Change Service Climate Data Store (https://doi.org/10.24381/cds.bd0915c6 and https://doi.org/10.24381/cds.adbb2d47). The original CERES SSF1deg products are available at CERES Data Products (https://ceres.larc.nasa.gov/data/#ssf1deg-level-3).

Acknowledgments

This material is based upon work supported by the U.S. Department of Energy (DOE), Office of Science, Office of Advanced Scientific Computing Research and Office of Biological and Environmental Research, Scientific Discovery through Advanced Computing (SciDAC) program under Award Number DE-SC0018650 and partially supported by the U.S. DOE, Office of Science, Office of Biological and Environmental Research, Atmospheric System Research (ASR) program Interagency Award 89243020SSC000055. The computational and storage resources are provided by the National Energy Research Scientific Computing Center (NERSC, supported by the Office of Science of the U.S. DOE under Contract No. DE-AC02-05CH11231) and the NOAA Research and Development High Performance Computing Program (https://rdhpcs.noaa.gov). We thank Hsiang-He Lee for discussions about the E3SM-FIVE code and Pornampai Narenpitak for sharing her version of the code for calculating EIS. The E3SM-FIVE code used for this research is available at https://github.com/E3SM-Project/E3SM/tree/hsiangheleellnl/atm/FIVE_LS_v2.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., & Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems*. Retrieved from https://www.tensorflow.org/

Barton, N. P., Klein, S. A., Boyle, J. S., & Zhang, Y. (2012). Arctic synoptic regimes: Comparing domain-wide Arctic cloud observations with CAM4 and CAM5 during similar dynamics. *Journal of Geophysical Research: Atmospheres*, *117*(D15205). https://doi.org/10.1029/2012JD017589

Bennhold, F., & Sherwood, S. (2008). Erroneous relationships among humidity and cloud forcing variables in three global climate models. *Journal of Climate*, *21*(17), 4190–4206. https://doi.org/10.1175/2008JCLI1969.1

Bogenschutz, P. A., Yamaguchi, T., & Lee, H.-H. (2021). The Energy Exascale Earth System Model simulations with high vertical resolution in the lower troposphere. *Journal of Advances in Modeling Earth Systems*, *13*(6), e2020MS002239. https://doi.org/10.1029/2020MS002239

Bony, S., Dufresne, J.-L., Le Treut, H., Morcrette, J.-J., & Senior, C. (2004). On dynamic and thermodynamic components of cloud changes. *Climate Dynamics*, *22*(2–3), 71–86. https://doi.org/10.1007/s00382-003-0369-6

Brunke, M. A., Ma, P.-L., Eyre, J. J. R., Rasch, P. J., Sorooshian, A., & Zeng, X. (2019). Subtropical marine low stratiform cloud deck spatial errors in the E3SMv1 Atmosphere Model. *Geophysical Research Letters*, *46*(21), 12598–12607. https://doi.org/10.1029/2019GL084747

E3SM Project. (2018). *Energy Exascale Earth system model (E3SM)*. https://doi.org/10.11578/E3SM/dc.20180418.36

Géron, A. (2019). *Hands-on machine learning with Scikit-learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media.

Gettelman, A., Liu, X., Ghan, S. J., Morrison, H., Park, S., Conley, A. J., et al. (2010). Global simulations of ice nucleation and ice supersaturation with an improved cloud scheme in the Community Atmosphere Model. *Journal of Geophysical Research*, *115*(D18). https://doi.org/10.1029/2009JD013797

Gettelman, A., Morrison, H., Santos, S., Bogenschutz, P., & Caldwell, P. M. (2015). Advanced two-moment bulk microphysics for global models. Part II: Global model solutions and aerosol–cloud interactions. *Journal of Climate*, *28*(3), 1288–1307. https://doi.org/10.1175/JCLI-D-14-00103.1

Golaz, J.-C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., et al. (2019). The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2089–2129. https://doi.org/10.1029/2018MS001603

Golaz, J.-C., Larson, V. E., & 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), 3540–3551. https://doi.org/10.1175/1520-0469(2002)059<3540:apbmfb>2.0.co;2

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. The MIT Press. Retrieved from http://www.deeplearningbook.org

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1026–1034). https://doi.org/10.1109/iccv.2015.123

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Iacono, M. J., Mlawer, E. J., Clough, S. A., & Morcrette, J.-J. (2000). Impact of an improved longwave radiation model, RRTM, on the energy budget and thermodynamic properties of the NCAR Community Climate Model, CCM3. *Journal of Geophysical Research: Atmospheres*, *105*(D11), 14873–14890. https://doi.org/10.1029/2000JD900091

Jakob, C., & Tselioudis, G. (2003). Objective identification of cloud regimes in the Tropical Western Pacific. *Geophysical Research Letters*, *30*(21). https://doi.org/10.1029/2003GL018367

Kawai, H., Koshiro, T., & Webb, M. J. (2017). Interpretation of factors controlling low cloud cover and low cloud feedback using a unified predictive index. *Journal of Climate*, *30*(22), 9119–9131. https://doi.org/10.1175/JCLI-D-16-0825.1

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv:1412.6980v9

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

Kubar, T. L., Waliser, D. E., & Li, J. (2011). Boundary layer and cloud structure controls on tropical low cloud cover using a-train satellite data and ecmwf analyses. *Journal of Climate*, *24*(1), 194–215. https://doi.org/10.1175/2010JCLI3702.1

Larson, V. E. (2017). CLUBB-SILHS: A Parameterization of Subgrid Variability in the Atmosphere. arXiv:1711.03675

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

L'Ecuyer, T. S., Hang, Y., Matus, A. V., & Wang, Z. (2019). Reassessing the effect of cloud type on Earth's energy balance in the age of active spaceborne observations. Part I: Top of atmosphere and surface. *Journal of Climate*, *32*(19), 6197–6217. https://doi.org/10.1175/JCLI-D-18-0753.1

Lee, H.-H., Bogenschutz, P. A., & Yamaguchi, T. (2021). The implementation of framework for improvement by vertical enhancement into Energy Exascale Earth System Model. *Journal of Advances in Modeling Earth Systems*, *13*(6), e2020MS002240. https://doi.org/10.1029/2020MS002240

Lohninger, H. (1999). *Teach/me data analysis*. Springer.

Lohninger, H. (2006). *Neural networks: Extrapolation*. Retrieved from http://www.vias.org/tmdatanaleng/cc_ann_extrapolation.html

Ma, H.-Y., Xie, S., Klein, S. A., Williams, K. D., Boyle, J. S., Bony, S., et al. (2014). On the correspondence between mean forecast errors and climate errors in CMIP5 models. *Journal of Climate*, *27*(4), 1781–1798. https://doi.org/10.1175/JCLI-D-13-00474.1

Mechem, D. B., Wittman, C. S., Miller, M. A., Yuter, S. E., & De Szoeke, S. P. (2018). Joint synoptic and cloud variability over the Northeast Atlantic near the Azores. *Journal of Applied Meteorology and Climatology*, *57*(6), 1273–1290. https://doi.org/10.1175/JAMC-D-17-0211.1

Medeiros, B., & Stevens, B. (2011). Revealing differences in GCM representations of low clouds. *Climate Dynamics*, *36*(1–2), 385–399. https://doi.org/10.1007/s00382-009-0694-5

Mieslinger, T., Horváth, Á., Buehler, S. A., & Sakradzija, M. (2019). The dependence of shallow cumulus macrophysical properties on large-scale meteorology as observed in ASTER imagery. *Journal of Geophysical Research: Atmospheres*, *124*(21), 11477–11505. https://doi.org/10.1029/2019JD030768

Minnis, P., Sun-Mack, S., Chen, Y., Chang, F.-L., Yost, C. R., Smith, W. L., & Xie, Y. (2020). CERES MODIS cloud product retrievals for Edition 4. Part I: Algorithm changes. *IEEE Transactions on Geoscience and Remote Sensing*, *59*, 2744–2780. https://doi.org/10.1109/TGRS.2020.3008866

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres*, *102*(D14), 16663–16682. https://doi.org/10.1029/97JD00237

Mülmenstädt, J., Lubin, D., Russell, L. M., & Vogelmann, A. M. (2012). Cloud properties over the North Slope of Alaska: Identifying the prevailing meteorological regimes. *Journal of Climate*, *25*(23), 8238–8258. https://doi.org/10.1175/JCLI-D-11-00636.1

Myers, T. A., & Norris, J. R. (2013). Observational evidence that enhanced subsidence reduces subtropical marine boundary layer cloudiness. *Journal of Climate*, *26*(19), 7507–7524. https://doi.org/10.1175/JCLI-D-12-00736.1

Nam, C. C., & Quaas, J. (2013). Geographically versus dynamically defined boundary layer cloud regimes and their use to evaluate general circulation model cloud parameterizations. *Geophysical Research Letters*, *40*(18), 4951–4956. https://doi.org/10.1002/grl.50945

Norris, J. R., & Weaver, C. P. (2001). Improved techniques for evaluating GCM cloudiness applied to the NCAR CCM3. *Journal of Climate*, *14*(12), 2540–2550. https://doi.org/10.1175/1520-0442(2001)014<2540:itfegc>2.0.co;2

Park, S., & Shin, J. (2019). Heuristic estimation of low-level cloud fraction over the globe based on a decoupling parameterization. *Atmospheric Chemistry and Physics*, *19*(8), 5635–5660. https://doi.org/10.5194/acp-19-5635-2019

Pincus, R., Platnick, S., Ackerman, S. A., Hemler, R. S., & Hofmann, R. J. P. (2012). Reconciling simulated and observed views of clouds: MODIS, ISCCP, and the limits of instrument simulators. *Journal of Climate*, *25*(13), 4699–4720. https://doi.org/10.1175/JCLI-D-11-00267.1

Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for activation functions. arXiv:1710.05941v2

Slingo, J. M. (1987). The development and verification of a cloud prediction scheme for the ECMWF model. *Quarterly Journal of the Royal Meteorological Society*, *113*(477), 899–927. https://doi.org/10.1002/qj.49711347710

Su, H., Jiang, J. H., Zhai, C., Perun, V. S., Shen, J. T., Del Genio, A., et al. (2013). Diagnosis of regime-dependent cloud simulation errors in CMIP5 models using "A-Train" satellite observations and reanalysis data. *Journal of Geophysical Research: Atmospheres*, *118*(7), 2762–2780. https://doi.org/10.1029/2012JD018575

Trepte, Q. Z., Minnis, P., Sun-Mack, S., Yost, C. R., Chen, Y., Jin, Z., et al. (2019). Global cloud detection for CERES Edition 4 using Terra and Aqua MODIS data. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(11), 9410–9449. https://doi.org/10.1109/TGRS.2019.2926620

Tselioudis, G., Zhang, Y., & Rossow, W. B. (2000). Cloud and radiation variations associated with northern midlatitude low and high sea level pressure regimes. *Journal of Climate*, *13*(2), 312–327. https://doi.org/10.1175/1520-0442(2000)013<0312:carvaw>2.0.co;2

Wall, C. J., Hartmann, D. L., & Ma, P.-L. (2017). Instantaneous linkages between clouds and large-scale meteorology over the Southern Ocean in observations and a climate model. *Journal of Climate*, *30*(23), 9455–9474. https://doi.org/10.1175/JCLI-D-17-0156.1

Weaver, C. P., & Ramanathan, V. (1997). Relationships between large-scale vertical velocity, static stability, and cloud radiative forcing over Northern Hemisphere extratropical oceans. *Journal of Climate*, *10*(11), 2871–2887. https://doi.org/10.1175/1520-0442(1997)010<2871:rblsvv>2.0.co;2

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

Yamaguchi, T., Feingold, G., & Larson, V. E. (2017). Framework for improvement by vertical enhancement: A simple approach to improve representation of low and high-level clouds in large-scale models. *Journal of Advances in Modeling Earth Systems*, *9*(1), 627–646. https://doi.org/10.1002/2016MS000815

Zhang, G. J., & McFarlane, N. A. (1995). Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Centre general circulation model. *Atmosphere-Ocean*, *33*(3), 407–446. https://doi.org/10.1080/07055900.1995.9649539

Zhang, Y., Xie, S., Lin, W., Klein, S. A., Zelinka, M., Ma, P.-L., et al. (2019). Evaluation of clouds in version 1 of the E3SM atmosphere model with satellite simulators. *Journal of Advances in Modeling Earth Systems*, *11*(5), 1253–1268. https://doi.org/10.1029/2018MS001562

Zimek, A., Schubert, E., & Kriegel, H.-P. (2012). A survey on unsupervised outlier detection in high-dimensional numerical data. *Statistical Analysis and Data Mining*, *5*(5), 363–387. https://doi.org/10.1002/sam.11161