

# A machine learning parameterization of clouds in a coarse-resolution climate model for unbiased radiation

Brian Henn<sup>1</sup>, Yakelyn R. Jauregui<sup>2</sup>, Spencer K. Clark<sup>1,3</sup>, Noah D. Brenowitz<sup>4</sup>,  
Jeremy McGibbon<sup>1</sup>, Oliver Watt-Meyer<sup>1</sup>, Andrew G. Pauling<sup>5</sup>, Christopher S.  
Bretherton<sup>1</sup>

<sup>1</sup>Allen Institute for Artificial Intelligence, Seattle, WA

<sup>2</sup>University of Washington, Seattle, WA

<sup>3</sup>NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ

<sup>4</sup>NVIDIA Corporation, Santa Clara, CA

<sup>5</sup>University of Otago, Dunedin, NZ

## Key Points:

- Fine-grid model clouds, when coarsened, can produce more accurate radiative fluxes than a coarse model's own parameterized clouds
- Machine learning can capture the coarsened fine-grid clouds as functions of coarse model state such as temperature, pressure, and humidity
- With appropriate post-processing, the machine-learned clouds also outperform the coarse model's clouds in terms of diagnostic radiation

## Abstract

Coarse-grid weather and climate models rely particularly on parameterizations of cloud fields, and coarse-grained cloud fields from a fine-grid reference model are a natural target for a machine-learned parameterization. We machine-learn the coarsened-fine cloud properties as a function of coarse-grid model state in each grid cell of NOAA’s FV3GFS global atmosphere model with 200 km grid spacing, trained using a 3 km fine-grid reference simulation with a modified version of FV3GFS. The ML outputs are coarsened-fine fractional cloud cover and liquid and ice cloud condensate mixing ratios, and the inputs are coarse model temperature, pressure, relative humidity, and ice cloud condensate. The predicted fields are skillful and unbiased, but somewhat under-dispersed, resulting in too many partially-cloudy model columns. When the predicted fields are applied diagnostically (offline) in FV3GFS’s radiation scheme, they lead to small biases in global-mean top-of-atmosphere (TOA) and surface radiative fluxes. An unbiased global-mean TOA net radiative flux is obtained by setting to zero any predicted cloud with grid-cell mean cloud fraction less than a threshold of 6.5%; this does not significantly degrade the ML prediction of cloud properties. The diagnostic, ML-derived radiative fluxes are far more accurate than those obtained with the existing cloud parameterization in the nudged coarse-grid model, as they leverage the accuracy of the fine-grid reference simulation’s cloud properties.

## Plain Language Summary

Weather and climate models typically use simplified means of predicting clouds, and these methods are particularly important for models that run at coarse resolution (200 km pixels) and relatively quickly. Machine learning is a natural way to improve upon the methods of predicting cloud. We first show that detailed cloud information from an accurate model run at much finer resolution (3 km pixels) can produce more accurate radiative energy fluxes at earth’s surface and top of atmosphere when used in the coarse model, instead of its own predicted clouds. We show that by training on the temperature, humidity, and other characteristics of the coarse model, machine learning can replicate much of this skill in predicting clouds, and provide better radiation estimates than traditional methods. Because solar and thermal radiation is sensitive to small amounts of cloud, we find that we need to remove small amounts erroneously predicted by machine learning to achieve the right amounts of radiation at earth’s surface and top of atmosphere. The machine-learning based approach could be used as a replacement for the current statistical methods of predicting cloud fields in fast, coarse resolution weather and climate models.

## 1 Introduction

Accurately representing clouds is a central challenge in climate modeling. Surface and atmospheric radiative heating and precipitation formation are all mediated by cloud processes. Cloud feedbacks on climate change are the largest driver of uncertainty in climate sensitivity to greenhouse gas increases (Caldwell et al., 2016). Many types of cloud are highly spatially inhomogeneous on the 25–200 km grid scale of typical global climate models. Expert-designed parameterizations (simplified representations) of this subgrid variability are used in model predictions of grid-mean radiation and precipitation. Because clouds have diverse, complex spatial structures, developing such subgrid parameterizations is as much art as science, blending physical insights, empirical relationships, and post-hoc calibration of uncertain parameters.

The rise of machine learning (ML, i.e., data-driven models) capabilities has fostered new approaches to improving parameterizations (Gentine et al., 2018). Examples include replacing computationally-intensive physical parameterizations with ML emulation (Krasnopolsky

et al., 2005, 2010; O’Gorman & Dwyer, 2018; Keller & Evans, 2019; Lagerquist et al., 2021; Perkins et al., 2023) and training ML against observations (McGibbon & Bretherton, 2019; Watt-Meyer et al., 2021; Chen et al., 2023) or more accurate and computationally intensive parameterizations (Chantry et al., 2021). ML parameterizations for coarse-grid models have been trained on coarsened (coarse-grained) outputs of fine-grid reference simulations, e.g. to predict the effect of the full suite of physics parameterizations (Rasp et al., 2018; Brenowitz & Bretherton, 2019; Yuval et al., 2021), or a column-wise correction to the coarse-grid model physics (Bretherton et al., 2022; Clark et al., 2022; Kwa et al., 2023). While using ML in coarse-grid models to correct physics tendencies of temperature and humidity can improve aspects of their simulated climates, clouds are often made worse because they are not among the ML target variables (Kwa et al., 2023), creating knock-on biases in surface and top-of-atmosphere radiative fluxes. This motivated us to use ML to also improve the simulated cloud distributions.

Grundner et al. (2022, 2023) and Chen et al. (2023) have developed ML parameterizations of fractional cloud cover trained on coarsened fine-grid output and observations, and they showed that these parameterizations can improve upon the skill of existing physically-based parameterizations. We extend such work to demonstrate that ML-predicted cloud statistics, including fractional cloud cover and ice and liquid cloud condensate mixing ratios, can improve the simulation of coarse model radiative fluxes, given careful attention to the vertical overlap of fractional cloud cover within grid columns. Given the role that clouds play in the atmosphere’s radiative balance, we consider radiative fluxes as a key criteria for evaluating ML cloud parameterizations. Evaluating and optimizing the skill of ML-predicted clouds in producing appropriate precipitation is another important aspect not considered here, because it requires a more sophisticated treatment of the subgrid distribution of clouds and precipitation and is tightly coupled to the cloud microphysics parameterization.

In this study, we seek to produce ML cloud fields in a coarse-grid model that are unbiased both in terms of cloud and radiative fluxes. For simplicity, we use a gridcell-local approach (i.e., making each prediction only with inputs and outputs from the local gridcell in vertical and horizontal space), as cloud properties within a grid cell should be describable in terms of its internal state. Section 2 describe the fine-grid reference dataset, the coarse-grid model we use to compute radiative fluxes, and the ML approach. Section 3.1 describes the coarsened-fine clouds and shows that when passed through the coarse model radiation parameterization with suitable vertical overlap assumptions, they produce unbiased radiative fluxes. Section 3.2 shows the performance of the ML cloud approach. Finally, Section 3.3 demonstrates the application of a simple post-processing step to the ML clouds to produce unbiased radiation. Section 4 discusses next steps toward online application of our ML approach during climate simulations and more sophisticated potential approaches for ML of subgrid cloud variability.

## 2 Data and Methods

Our study uses three models: 1) a fine-grid global storm-resolving model to produce a reference dataset of clouds, radiation, and atmospheric state predictors such as temperature and humidity; 2) a coarse-grid, economical version of this global atmospheric model with a radiation parameterization that computes fluxes and heating rates given a coarse-grid representation of the cloud state; and 3) an ML model trained on the fine-grid cloud distribution that diagnoses cloud fields from coarse-grid atmospheric predictors; this is designed to replace the physical parameterizations used to predict cloud properties needed for the radiation parameterization in the coarse-grid model.

## 2.1 Atmospheric models

### 2.1.1 *Fine-grid reference model*

Our fine-grid reference model is X-SHiELD (Harris et al., 2020), a non-hydrostatic global atmosphere model with approximately 3 km horizontal grid spacing, developed by the NOAA Geophysical Fluid Dynamics Laboratory (GFDL). It uses a C3072 cubed-sphere grid and a hybrid pressure-sigma vertical coordinate. X-SHiELD shares the same FV3 dynamical core (Zhou et al., 2019) and most of its physics parameterizations with NOAA’s Global Forecast System (GFS), NOAA’s operational global weather forecast model.

X-SHiELD uses the GFDL microphysics scheme (Zhou et al., 2022), which performs inline microphysical moisture adjustments in the dynamical core, and the RRTMG radiation scheme (Mlawer et al., 2016) as implemented in GFS (Liu & Yang, 2023). For computational efficiency, RRTMG uses the Monte Carlo independent column approximation (MCICA, Pincus et al., 2003), which makes an unbiased, stochastic approximation to full shortwave and longwave radiation calculations within each grid column using a random sample of the cloud overlap configuration in each spectral band. RRTMG’s primary inputs are the fractional cloud cover and liquid and ice cloud condensate mixing ratios for each model cell (which we will predict using ML), along with ancillary features such as aerosol concentrations.

Our version of X-SHiELD was configured similarly to the year-long reference simulations used in Kwa et al. (2023) and Cheng et al. (2022). We made the following configuration changes relative to those simulations, in order to ensure the compatibility of the fine- and coarse-grid radiation schemes:

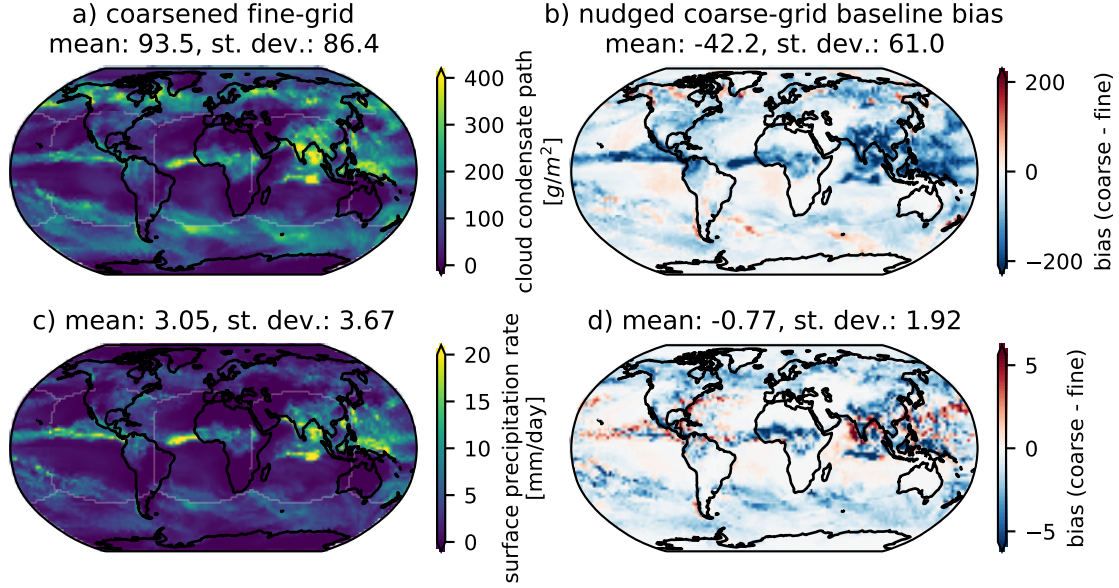
- The “ccnorm” namelist parameter is set to true (see Section 3.1)
- Vertical cloud overlap uses a latitude-dependent decorrelation length assumption
- The “cloud\_gfdl” and “pdfcld” namelist parameters are turned off
- The radiation scheme was run every 900 s, instead of every 1800 s.

A ten day X-SHiELD simulation on 79 vertical levels was initialized from a set of restart files from the simulation used in Kwa et al. (2023) at 00 UTC on 31 July 2020, using a physics time step of 180 s and 40 dynamical sub-steps per physics time step to ensure model numerical fidelity.

We use hourly outputs of the simulated model state and derived diagnostics. Given the size of a global field of data on a 3 km, 79-level grid, we implemented online coarse-graining of the model state and diagnostics following Bretherton et al. (2022). The outputs needed from the fine-grid model were horizontally coarsened by a factor of 64 to 200 km resolution before being stored. Two-dimensional variables (e.g. surface and TOA radiative fluxes) were coarse-grained using area-weighted horizontal averaging while three-dimensional variables on vertical model levels (e.g. liquid and ice cloud condensate mixing ratios, fractional cloud cover, and thermodynamic fields such as radiative heating rates) were coarsened along the coarse grid’s spatially and temporally varying pressure levels.

### 2.1.2 *Nudged coarse-grid baseline model*

The coarsened temperature, humidity, winds, and pressure thicknesses from the fine-grid simulation were used to nudge a 200 km grid version of FV3GFS with the same 79 vertical model levels, following Bretherton et al. (2022). FV3GFS (Putman & Lin, 2007) combines the GFS physical parameterizations with the FV3 dynamical core; it shares much of the same model code and physical parameterizations with X-SHiELD (Zhou et al., 2019). One important exception is that the GFS deep convection scheme is active in the coarse FV3GFS simulation, but not in X-SHiELD, which resolves individual cu-



**Figure 1.** Coarsened-fine fields related to clouds (a, c) and their biases in the coarse-grid model (b, d) that is nudged to the coarsened fine-grid temperature, humidity, winds, and pressure thickness. Day 8 through 10 time-mean fields are shown and used to compute the statistics.

mulus updrafts. To the extent possible, we run FV3GFS with the same GFDL cloud microphysics and RRTMG radiation parameterization configurations as in X-SHiELD.

When running the coarse-grid FV3GFS, we nudge its prognostic ‘memory’ variables at each time (temperature, specific humidity, horizontal winds, and layer pressure thickness) to the coarsened fine-grid state, using a 3-hour nudging timescale, following Bretherton et al. (2022). This ensures that the coarse model state evolves in an internally consistent way that remains very close to that of the fine-grid reference, allowing for meaningful comparison of their clouds and radiation fields.

One might hope that nudging of the prognostic memory variables would also ensure that the nudged coarse simulation produces clouds and radiation that are similar to the coarsened-fine reference, but this is not the case (Bretherton et al., 2022). Fig. 1 shows that the coarse nudged FV3GFS simulation has large negative biases in both cloud condensate path and surface precipitation rate. That is, the coarse model physics parameterizations produce significantly less cloud and precipitation than the coarsened-fine reference for the same column profiles of temperature and humidity. Because the nudged simulation has insufficient cloud, it also has excessive longwave and shortwave radiative transmissivity, as shown in Section 3.1.

### 2.1.3 Diagnostic radiation scheme

To compute the radiative fluxes arising from coarse-grid cloud fields (both coarsened from the fine-grid model and predicted by ML), we use an offline implementation of the RRTMG radiation scheme used in X-SHiELD and FV3GFS. This version of RRTMG has been rewritten in Python and validated against its original Fortran implementation in terms of surface and top of atmosphere (TOA) flux accuracy (see Supporting Information S1). With the offline RRTMG scheme coupled to FV3GFS, fractional cloud cover and liquid and ice cloud condensate mixing ratios can be prescribed, while other RRTMG

input variables (temperature, humidity, land surface information, aerosols, etc.) are taken from the coarse model state.

## 2.2 Machine learning of coarsened fine-grid clouds

Our ML approach predicts coarsened fine-grid values of the three cloud properties needed by the RRTMG radiation scheme (fractional cloud cover and liquid and ice cloud condensate mixing ratios), based on the coarse-grid state. The ML is trained to best match these coarsened-fine cloud properties. Our hope is that if these ML cloud properties are used as inputs to the radiation scheme, the shortwave and longwave surface and TOA radiative fluxes will also be close to the coarsened-fine radiative fluxes, at least by comparison with using the parameterized clouds generated by the nudged coarse model, for the same column temperature and moisture profile. Because we lack the ability to back-propagate parameter gradients through our current version of the RRTMG radiation scheme, we train ML with the cloud properties as the target, and use slight additional post-processing to ensure nearly unbiased radiative fluxes without degrading the cloud predictions. This approach also provides physically interpretable cloud outputs.

For simplicity, our ML uses cell-local input features, since cloud can rapidly adjust through condensation, evaporation and precipitation to a changing local environment. We acknowledge that this assumption neglects the learnability of vertical cloud overlap between nearby grid cells in the same grid column, as well as possible learnable impacts of grid-nonlocal processes such as cumulus updrafts and downdrafts, which can cause large subgrid cloud inhomogeneity within grid cells. We also recognize that a more complex ML cloud parameterization could predict the full subgrid distributions of cloud liquid and ice condensate, information that might be used to make the radiative fluxes more accurate and to learn precipitation fluxes. Nevertheless, our cell-local ML approach works well enough for the purposes of this paper.

We use a neural network (NN) with a fully-connected (dense) multi-layer perceptron architecture (Hastie et al., 2009). The NN input features are coarse model grid-cell air temperature, relative humidity, and pressure, as well as the cloud ice mixing ratio produced by the coarse model’s physics. The last feature is included because cumulus updrafts, which are much smaller than the coarse grid scale, are a major source of cloud ice, and we find that cloud ice predictions are better when the coarse model physics parameterization’s output is included than when predicting from coarsened-fine thermodynamic properties of the coarse grid cell alone (not shown). The features are obtained from the coarse model alone, allowing this approach to be applied to improve the representation of clouds and their radiative effects in free-running coarse model simulations, something which has not been possible with previous cloud ML parameterizations, (e.g., Grundner et al., 2022). However, in this initial study we evaluate the ML skill only diagnostically, a necessary but much easier first step toward prognostic implementation.

Training data consists of sets of three-dimensional input and output variables from each hour during the first seven days of the 10-day X-SHiELD run and the corresponding 10-day nudged coarse FV3GFS run, after discarding an initial 6-hour spin-up period. This results in  $1.8 \times 10^8$  training samples. Validation was done on hourly output from the last three days of the reference simulation, a total of  $7.9 \times 10^7$  samples.

The NN is optimized based on hyperparameter sweeps. It has three hidden layers, each with a width of 169 neurons (a total of  $5.9 \times 10^4$  free parameters). Mean squared error loss summed over the three outputs is used, with the targets standard-normalized based on  $5 \times 10^5$  randomized samples first before computing loss. An additional layer is added to the NN to prevent its final outputs from leaving specified ranges during training and prediction; for fractional cloud coverage the range is  $[0, 1]$ , and for the mixing ratios the range is  $\geq 0$ . The NN is trained with stochastic gradient descent (SGD) using the Adam optimizer (Kingma & Ba, 2014) for 20 epochs with an exponential decay



learning rate schedule (initial rate  $10^{-3}$  and decay of 0.96 every  $10^5$  SGD steps). 512 samples per SGD batch were used, resulting in approximately  $6.9 \times 10^6$  steps.

We train four NNs with different random seeds. We select the one that has the smallest global-mean TOA net radiation bias and individual shortwave and longwave bias magnitudes on the validation dataset, as compared to radiative fluxes resulting from the coarsened-fine cloud. The other seeds have similar skill and their global-mean TOA net radiation scatters over a 2-3  $\text{W/m}^2$  range.

### 3 Results

We first describe the coarsened fine-grid cloud fields and radiative fluxes, since the cloud fields are the ML target and the radiative fluxes are what we ultimately wish to match. We next test whether prescribing the coarsened-fine clouds in a nudged coarse model run in place of the parameterized clouds removes most of its radiation biases, relative to the coarsened-fine output. Then we describe the skill of the ML-derived cloud fields in predicting the target coarsened-fine clouds and radiation fields. Lastly, we show how to achieve unbiased global-mean net TOA radiation via a post-processing step of thresholding of small predicted cloud fractions.

#### 3.1 Coarsened-fine clouds and resulting radiative fluxes

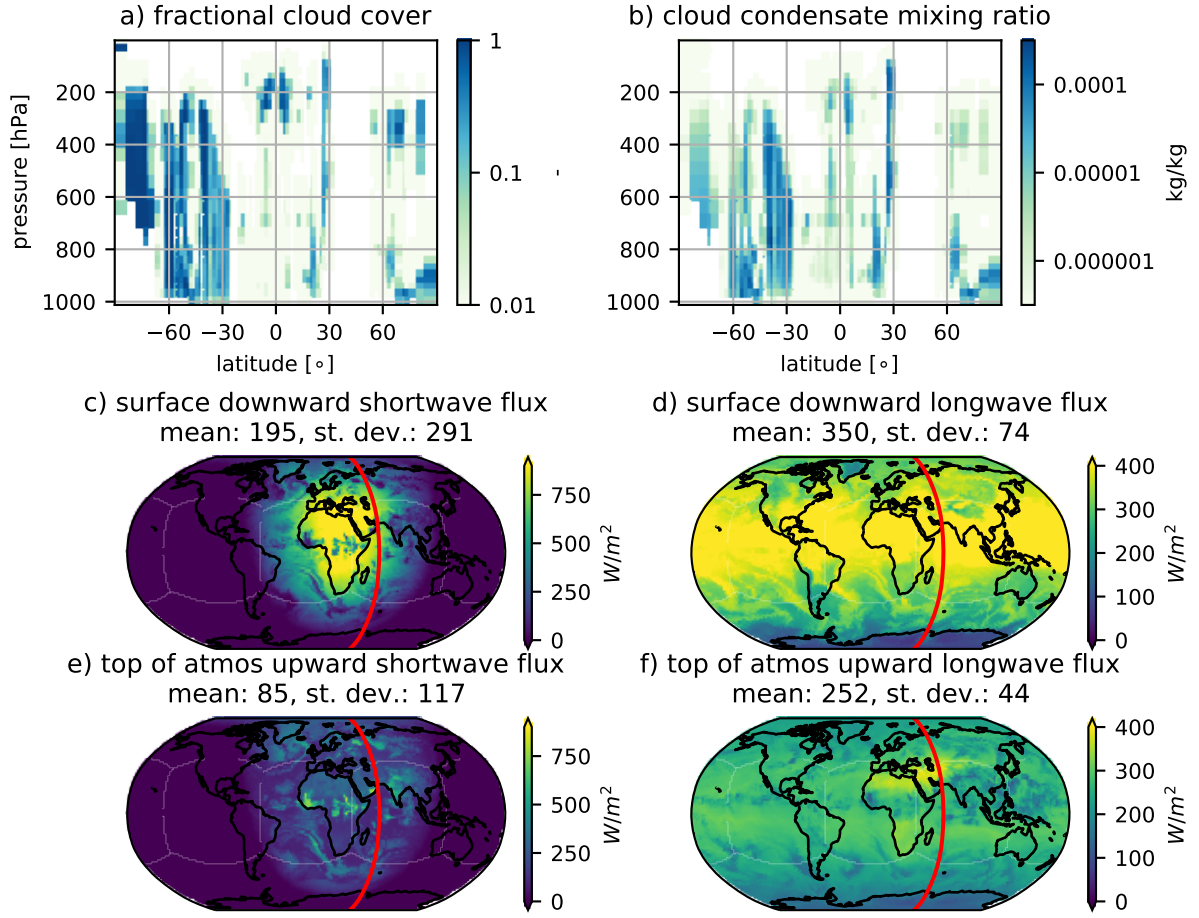
Fig. 2 shows that the coarsened-fine fractional cloud cover and grid-mean condensate fields exhibit a great deal of structure, including clear columns and ones with deep, extensive cloud. Condensate spans several orders of magnitude. The radiation fields exhibit correspondingly detailed spatial variability, particularly for shortwave. An ML cloud parameterization needs to be able to capture these features.

We compute the radiative fluxes obtained by prescribing the coarsened-fine cloud fields in the nudged coarse FV3GFS simulation. This is purely diagnostic; these do not feed back into the nudged coarse model evolution. Because of the nudging, the coarse model temperature and humidity profiles are nearly identical to their coarsened-fine counterparts, although the cloud amounts are not.

Fig. 3 shows maps of the time-mean biases in surface and TOA shortwave and longwave radiative fluxes, for the prescribed coarse-grained fine-resolution cloud fields and for the nudged coarse model's own cloud fields, the baseline upon which we aim to improve. The nudged coarse model's lack of cloud condensate translates into substantial global-mean radiation flux biases and even larger regional biases. By prescribing the coarsened-fine clouds, we achieve nearly unbiased radiative fluxes, with time-mean spatial pattern error magnitudes reduced by 30-60% for longwave radiation and around 80% for shortwave radiation. Thus, if ML could skillfully emulate the coarsened-fine cloud fields, it would also greatly improve the coarse model radiation fields.

Achieving unbiased radiative fluxes for the prescribed coarsened-fine cloud runs required consistency between the coarse and fine model radiation scheme in the choice of cloud overlap assumption and the setting of an FV3GFS namelist parameter ("cnorm", discussed further below) governing condensate scaling in fractionally-cloud covered cells. Table 1 shows the bias for different choices, with the last column being the parameter choices that produce nearly unbiased radiative fluxes shown in Fig. 3. Absolute global-mean values exceeding 5  $\text{W/m}^2$  constitute large biases for climate modeling purposes, larger than the radiative effect of doubling  $\text{CO}_2$ . By this measure, the radiative biases are evidently sensitive to these choices.

While coarse-grid and (to a lesser extent) fine-grid simulations are sensitive to these radiation scheme settings, these sensitivities are more pronounced when using coarsened-fine cloud properties in the radiation scheme. The reason for this can be inferred from

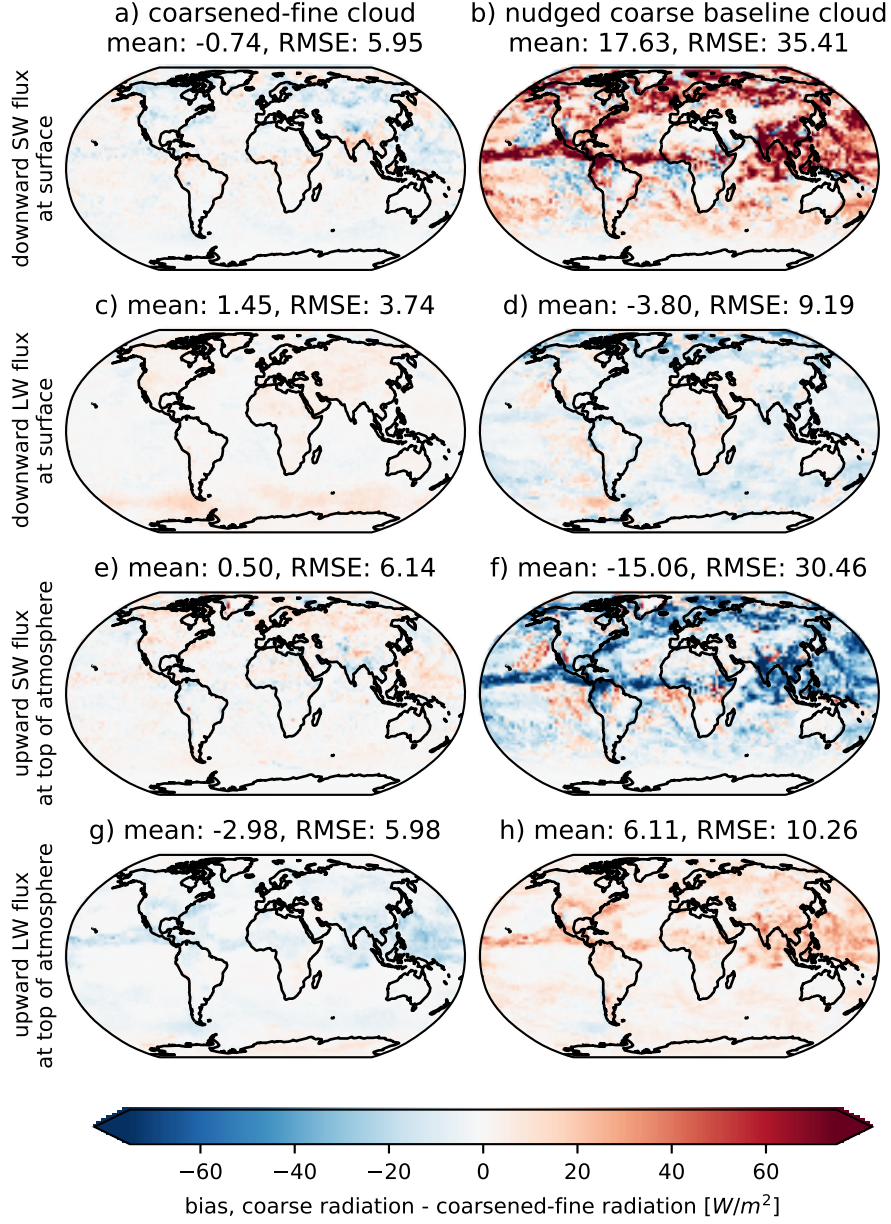


**Figure 2.** Coarsened-fine fractional cloud cover and condensate transects (a, b) and 15-minute average coarsened-fine radiative fluxes fields (c, d, e, f) for a typical simulated time, 1030 UTC on the seventh day of the fine-grid reference simulation. The transects are shown through the red line at 60°E.

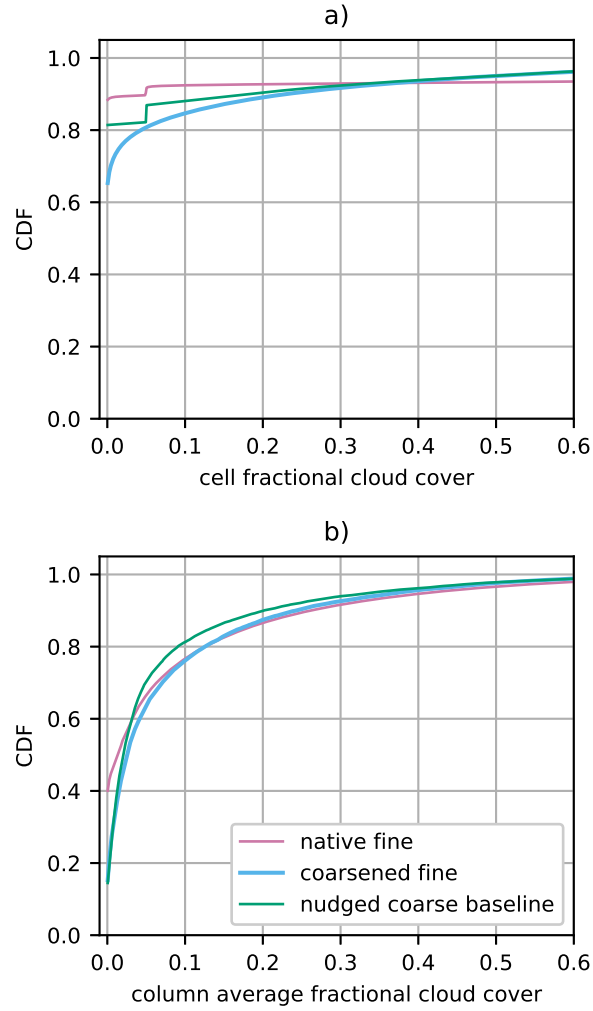
Bias [ $W/m^2$ ]	Cloud dataset					
	ccnorm=False			ccnorm=True		
	max-random	random	decorrelation	max-random	random	decorrelation
SW down at sfc.	16.30	8.80	13.86	4.69	-16.90	-0.74
LW down at sfc.	-2.84	2.54	-1.50	-0.70	10.08	1.45
SW up at TOA	-13.82	-7.57	-11.77	-4.05	14.14	0.50
LW up at TOA	2.28	-0.61	1.31	-1.06	-7.51	-2.98

**Table 1.** Global-mean surface and TOA radiative flux biases (over days 8 through 10) when the coarsened-fine cloud fields are prescribed in the coarse-grid FV3GFS model. *max-random*, *random* and *decorrelation* are three types of cloud overlap assumptions, and *ccnorm* describes how fractional cloud coverage is handled.

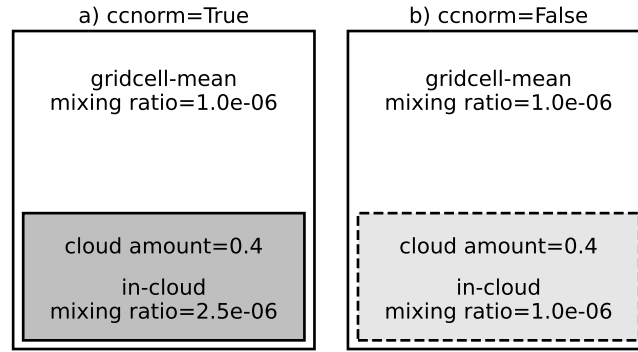




**Figure 3.** Surface and TOA radiative flux biases for the prescribed coarsened-fine cloud fields (a, c, e, g) and the baseline coarse nudged cloud fields (b, d, f, h). Bias and root-mean-squared error (RMSE) are computed over days 8 through 10 of the simulation period.



**Figure 4.** Cumulative distribution functions (CDFs) of the fractional cloud cover for individual model cells (a) and for the average over model columns (b), for a snapshot of the native fine, coarse-grained fine and coarse resolution datasets.



**Figure 5.** Schematic description of the effect of the “ccnorm” parameter in RRTMG. When set to true (a), a physically-conservative cloud condensate mixing ratio is used in computing radiative transmissivity. When set to false (b), the same condensate mixing ratio is used in the cloud as in the gridcell-mean, even if the cell is partially cloudy.

the distributions of fractional cloud cover in the different cloud datasets (Fig. 4a). Around 90% of fine-grid model cells have fractional cloud cover of zero, and most of the rest have nearly 100% fractional cloud cover. However, as an inevitable result of horizontal coarsening, the coarsened-fine dataset has less than 70% clear cells, with 20% of cells having fractional cloud cover between zero and 0.2. This makes the radiation scheme particularly sensitive to cloud overlap and subgrid condensate partitioning assumptions for the coarsened-fine data. For instance, less than 20% of the coarse grid columns have coarsened-fine fractional cloud cover less than  $10^{-3}$ , whereas about 40% of fine-grid columns do (Fig. 4b).

The “ccnorm” namelist parameter affects the subgrid partitioning of grid-mean condensate within fractionally-cloudy cells in RRTMG. Its effect is shown in Fig. 5. When this logical flag is set to true, RRTMG correctly scales up the in-cloud condensate mixing ratios, representing the given grid-mean condensate amount and the given fractional cloud cover in a physically conservative manner. When it is off, the in-cloud condensate is specified to be equal to the grid-mean condensate, regardless of the cloud fraction, which is physically incorrect.

The existence of the “ccnorm” flag may derive from the GFS implementation of the stochastic MCICA parameterization in RRTMG, which makes the correct choice of “ccnorm” less intuitively obvious. In particular, for each radiation band, a random number uniformly distributed between zero and one is generated, and for that band, the grid cell is taken to be fully cloud filled when the random number is less than the cloud fraction, and cloud-free when that random number is greater than the cloud fraction. It is easy to fall into the trap of then setting the condensate amount equal to the grid-mean condensate amount when that cell is chosen to be fully cloud filled, which corresponds to “ccnorm” being false. Indeed, “ccnorm” is false by default in the GFS RRTMG scheme of both X-SHIELD and the current version of NCEP’s operational global weather forecast model. Fig. 5 shows that this is not the physically correct approach, even in an operational setting where it may be compensating for other parameterization biases. The random number should be construed as drawing from a random sub-region of the coarse grid cell when doing the MCICA radiation calculation. If that random sub-region is cloudy, the radiation should be calculated using the in-cloud condensate (Fig. 5a), not the grid-

mean condensate (Fig. 5b). This is particularly important for cells with small but nonzero cloud fraction, for which the in-cloud condensate is much larger than the grid-mean condensate.

Setting “ccnorm” to true results in RRTMG seeing greater cloud optical depth, producing more reflected TOA shortwave and less outgoing TOA longwave radiation, which better matches the coarsened-fine reference radiation (Table 1). The global-mean increase in TOA upwelling shortwave radiation is particularly large: 10–22 W/m<sup>2</sup> depending on the chosen vertical cloud overlap scheme (see below). TOA upwelling longwave radiation is reduced by 3–7 W/m<sup>2</sup> in the global mean. Corresponding changes are seen in global-mean downwelling shortwave and longwave radiation at the surface.

For consistency, the 10-day fine-grid simulation used here was also run with “ccnorm” set to true. Over most parts of the world, the fine-grid radiative fluxes are only weakly affected by this choice, because fine-grid cells tend to be either clear or nearly entirely cloud-filled. One exception is over the Southern Ocean, where small subgrid cloud fractions are often generated by the shallow cumulus parameterization, even on the fine grid. There, setting “ccnorm” to true increases the regional time-mean reflected shortwave radiation simulated by the fine-grid model by several W/m<sup>2</sup> (not shown).

The radiative fluxes are also sensitive to the cloud overlap assumption. Two commonly used methods are maximum-random overlap (the GFS RRTMG default), in which subgrid cloud in contiguous vertical layers is assumed to overlap as much as possible, and random overlap, in which the horizontal distribution of subgrid clouds is assumed to be uncorrelated between vertical layers. A third method available in the FV3GFS RRTMG implementation, decorrelation overlap, assumes that the subgrid distribution of clouds within nearby vertical levels is correlated, with a correlation coefficient that decays exponentially with an empirically-specified, latitude-dependent  $e$ -folding scale. This option is arguably most physically realistic.

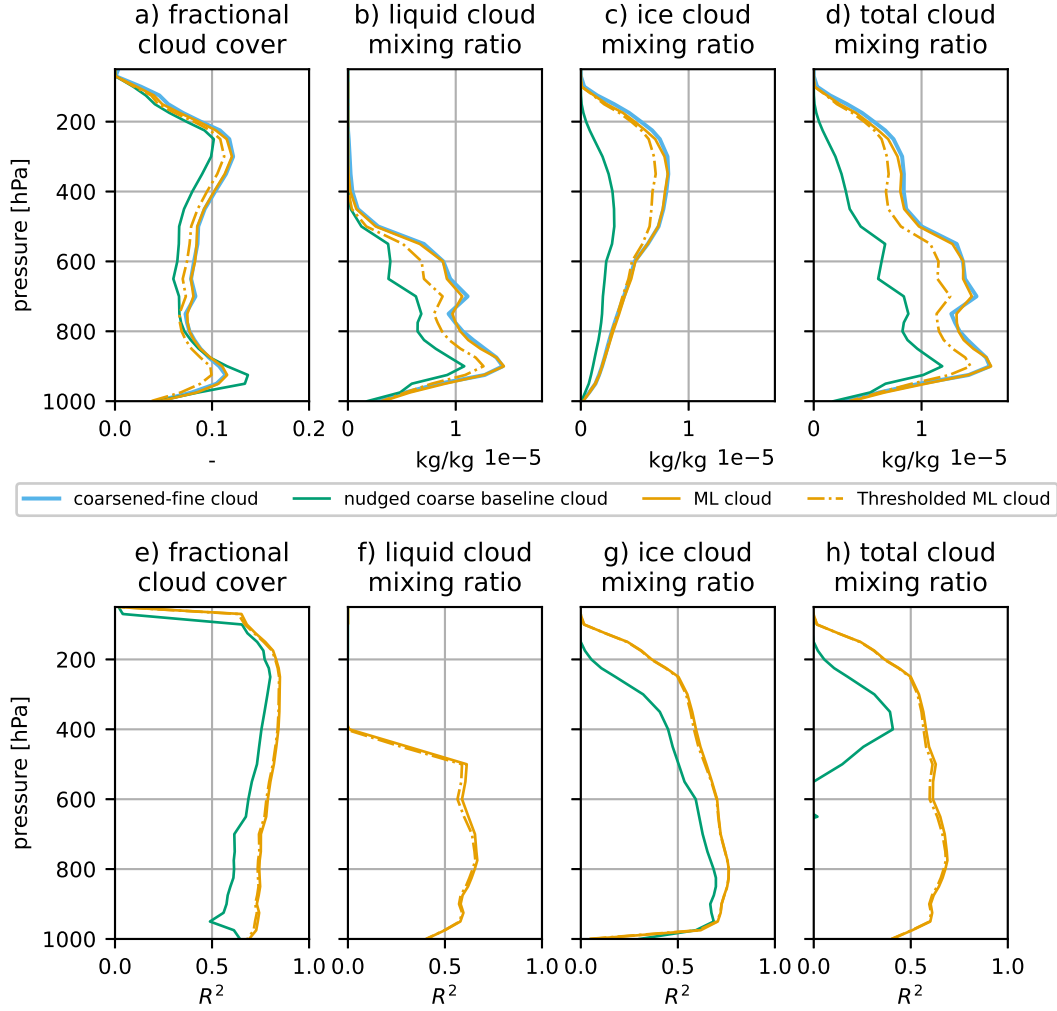
For the fine-grid model, the TOA and surface fluxes are relatively insensitive to the choice of overlap, because the majority of fine-grid cells are clear or mostly cloud filled, for which the spatial distribution of subgrid cloud is a moot point. However, for the coarsened-fine output, the overlap scheme has a large impact, as seen in Table 1. We selected decorrelation overlap because it is physically attractive and gave minimal global-mean TOA shortwave and longwave biases.

### 3.2 ML clouds

Since the coarsened-fine cloud field can produce nearly unbiased radiative fluxes, an ML approximation of those cloud properties might also be able to do so. The design principles and implementation of the simple ML scheme that we used for this purpose was detailed in Sec. 2.2; here we show results for the best-performing NN seed ensemble member.

The solid orange lines in Figs. 6a–c show that the three ML predictands (fractional cloud cover and liquid and ice cloud condensate mixing ratios) have vertical profiles with relatively unbiased global means relative to the coarsened-fine validation data. Fig. 6d shows that this also holds for cloud condensate (the sum of cloud liquid and ice). In contrast, the nudged coarse baseline cloud fields have large negative condensate biases, and negative fractional cloud cover biases throughout most of the troposphere.

All the ML cloud properties have an  $R^2$  between 0.5 and 0.7 between the surface and 200 hPa (Fig. 6e–h). This skill degrades near and above the tropopause, but magnitudes of fractional cloud cover and condensate at these levels are very small. The skill of the ML-predicted cloud fields exceeds that of the nudged coarse baseline cloud fields throughout the atmospheric column.



**Figure 6.** Global- and time-mean vertical profiles over the validation period of fractional cloud cover (a) and liquid, ice, and total cloud condensate mixing ratios (b, c, d) for the coarsened-fine, coarse nudged baseline, and ML cloud datasets. e-h) Vertical profiles of  $R^2$  of the instantaneous fields, computed at each level using the global horizontal area-weighted mean for that variable. The  $R^2$  of the nudged coarse baseline is below 0 everywhere for liquid cloud mixing ratio (f). For both bias and  $R^2$ , the thresholded ML cloud dataset is also shown; see Section 3.3.

Fig. 7a-b shows that the ML overpredicts the occurrence of cells with small fractional cloud cover and small cloud condensate mixing ratio. We attribute this to coarse-grid thermodynamic states in which cloud may or may not be present, so the ML predictors must average over those conditions, while underpredicting the fraction of cloud-free cells. Fig. 7c compares the cumulative distribution function (CDF) of the resulting the column-integrated cloud condensate path for the ML predictions and the training data, which is a good proxy for the cloud impact on TOA and surface radiative fluxes, especially in the shortwave band. The solid line at  $10^{-3}$  kg/m<sup>2</sup> is a rough threshold for a radiatively significant condensate path. The fractional cloud cover and condensate CDF biases translate into too many columns with ML-predicted condensate path between  $10^{-3}$ - $10^{-1}$  kg/m<sup>2</sup> compared to the training dataset, in which about 20% of all columns are below this value. We infer that the ML overpredicts the fraction of grid columns with radiatively significant cloud. This suggests that the ML will also overpredict TOA reflected shortwave radiation and (because these condensate biases also apply to upper-tropospheric cirrus clouds) underpredict TOA outgoing longwave radiation.

Geographically, this bias translates into the ML overpredicting fractional cloud cover and cloud condensate path in regions of the globe with very little cloud, such as in the dry subtropics adjacent to regions of convection. This can be seen in the typical snapshot shown in Fig. 8. On the other hand, Fig. 7c shows that the ML slightly underpredicts the frequency of the highest condensate paths; geographically this leads to the ML underestimating cloud condensate maxima in areas of deep convection and strong frontal convergence.

The ML bias in underpredicting cloud-free columns is disproportionately important to global-mean radiative fluxes, given the small amounts of spurious ML-predicted cloud condensate in these columns. For example, the RRTMG scheme computes cloudy cell radiative extinction coefficient from liquid water  $\tau_{liq}$  as a function of liquid water path, *LWP* (Liu and Yang (2023), following parameters from Hu and Stamnes (1993)):

$$\tau_{liq} = LWP \cdot \left( \frac{a}{r_e^b} + c \right) \quad (1)$$

where  $r_e$  is an effective radius of cloud droplets, and  $a$ ,  $b$ , and  $c$  are semi-physical fitted parameters. For typical values of  $r_e$  (10  $\mu$ m) and the fitted parameters ( $a \approx 1800$ ,  $b \approx 1.1$ , and  $c \approx 8.0$  for shortwave at 750 nm), if the cloud liquid condensate mixing ratio is  $10^{-6}$  kg/kg over a cloud 10 hPa thick, then  $LWP=10^{-4}$  kg/m<sup>2</sup> and  $\tau_{liq}=0.015$ , and this will reduce the shortwave transmissivity in the cloud by 1.5% below clear-sky values. Cell condensate mixing ratios of  $10^{-6}$  kg/kg over a significant portion of the column and column-integrated condensate path approaching  $10^{-3}$  kg/m<sup>2</sup> are thus “radiatively significant” thresholds and are highlighted on Fig. 7b-c.

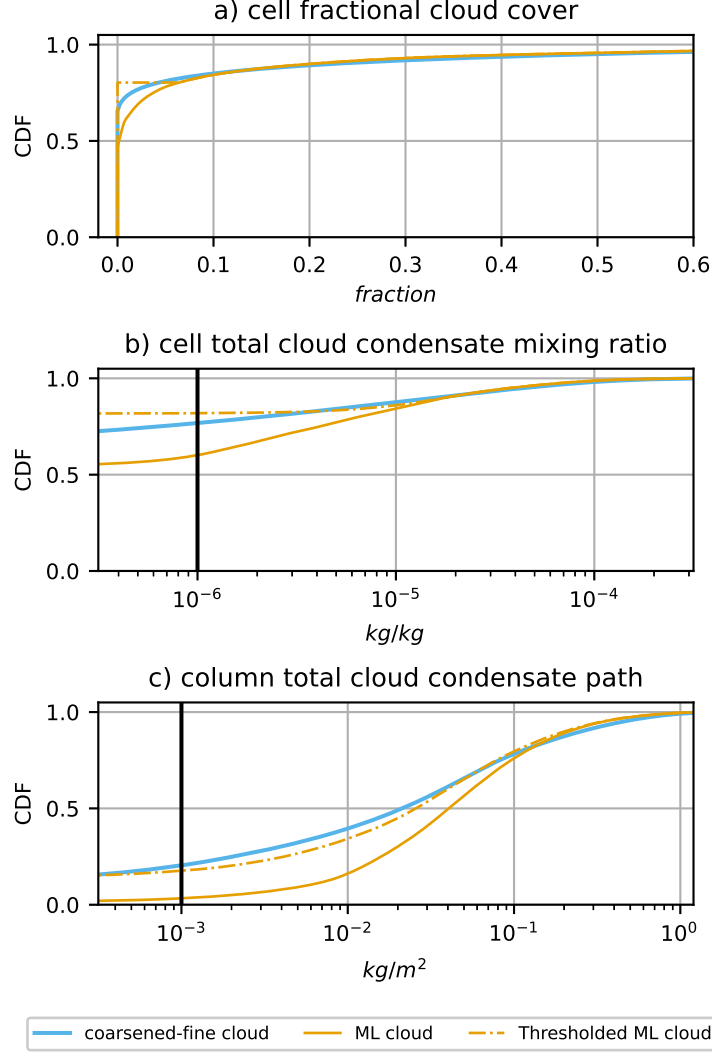
While the ML cloud model makes skillful and unbiased predictions, the radiative fluxes resulting from those predictions have significant global-mean biases, when compared to the coarsened-fine clouds’ radiative fluxes. Table 2 shows that with ML clouds, the TOA and surface radiative fluxes have global-mean biases that are 30-52% of the coarse nudged baseline, but of opposite sign. This is suggestive of excessive cloud optical depth, yet ML-predicted cloud condensate amounts are unbiased throughout the troposphere.

These radiative biases are an inevitable consequence of the ML making unbiased but under-dispersed predictions of cloud condensate and amount. To illustrate this, we define a normalized surface downward shortwave cloud radiative effect:

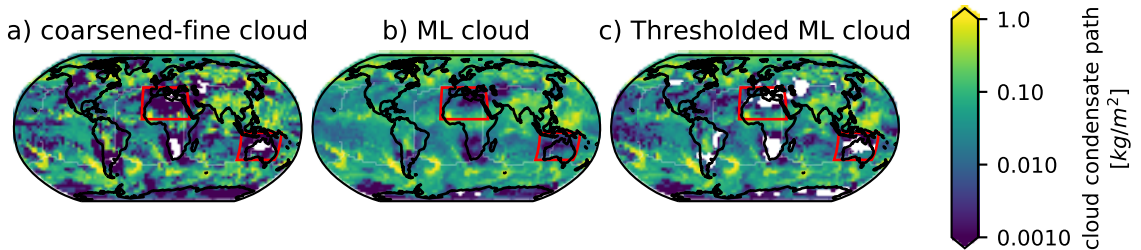
$$NSWCRE^{sfc} = \frac{SW_{clear-sky}^{\downarrow, sfc} - SW_{total-sky}^{\downarrow, sfc}}{SW_{clear-sky}^{\downarrow, sfc}} \quad (2)$$

where  $\downarrow, sfc$  indicates downward flux at the surface, and clear-sky and total-sky are the RRTMG scheme’s fluxes without and with cloud effects. This quantity is defined only for columns in which TOA downward shortwave flux is non-zero. In contrast to typical





**Figure 7.** CDFs of fractional cloud cover (a), total cloud condensate mixing ratio (b) and column-integrated total condensate path (c) for coarsened-fine and ML cloud. The thresholded ML cloud is also shown. “Radiatively significant” (see text) values for cell total condensate mixing ratio and column total condensate path are shown with vertical black lines.



**Figure 8.** Column-integrated cloud condensate path maps for a snapshot in the validation period, for the coarsened-fine target (a), the ML cloud predictions (b), and the thresholded ML cloud (c). Red boxes highlight regions where the predictions tend to overestimate the extent of thin cloud.

Metric	Cloud dataset			
		coarse nudged baseline	ML cloud	Thresholded ML cloud
Validation bias [ $\text{W/m}^2$ ]	SW down at sfc.	18.37	-6.15	-1.53
	LW down at sfc.	-5.24	2.74	1.40
	SW up at TOA	-15.56	4.80	0.86
	LW up at TOA	9.09	-4.33	-2.77
$R^2$ of validation time-mean	SW down at sfc.	0.840	0.956	0.959
	LW down at sfc.	0.980	0.992	0.993
	SW up at TOA	0.594	0.895	0.899
	LW up at TOA	0.839	0.940	0.949

**Table 2.** Bias and  $R^2$  of radiative fluxes from the coarse nudged baseline dataset, the ML cloud dataset, and the thresholded ML (section 3.3). Metrics are computed with reference to the radiative fluxes produced by the coarse-grained fine-resolution cloud fields, over hourly data from the ML validation period (days 8 through 10 of the reference simulation).

cloud radiative effect, here  $\text{NSWCRE}^{sfc}$  is the fraction of the clear-sky shortwave flux that is prevented from reaching the surface by clouds. A similar quantity could be defined for upward shortwave at TOA, with similar results.

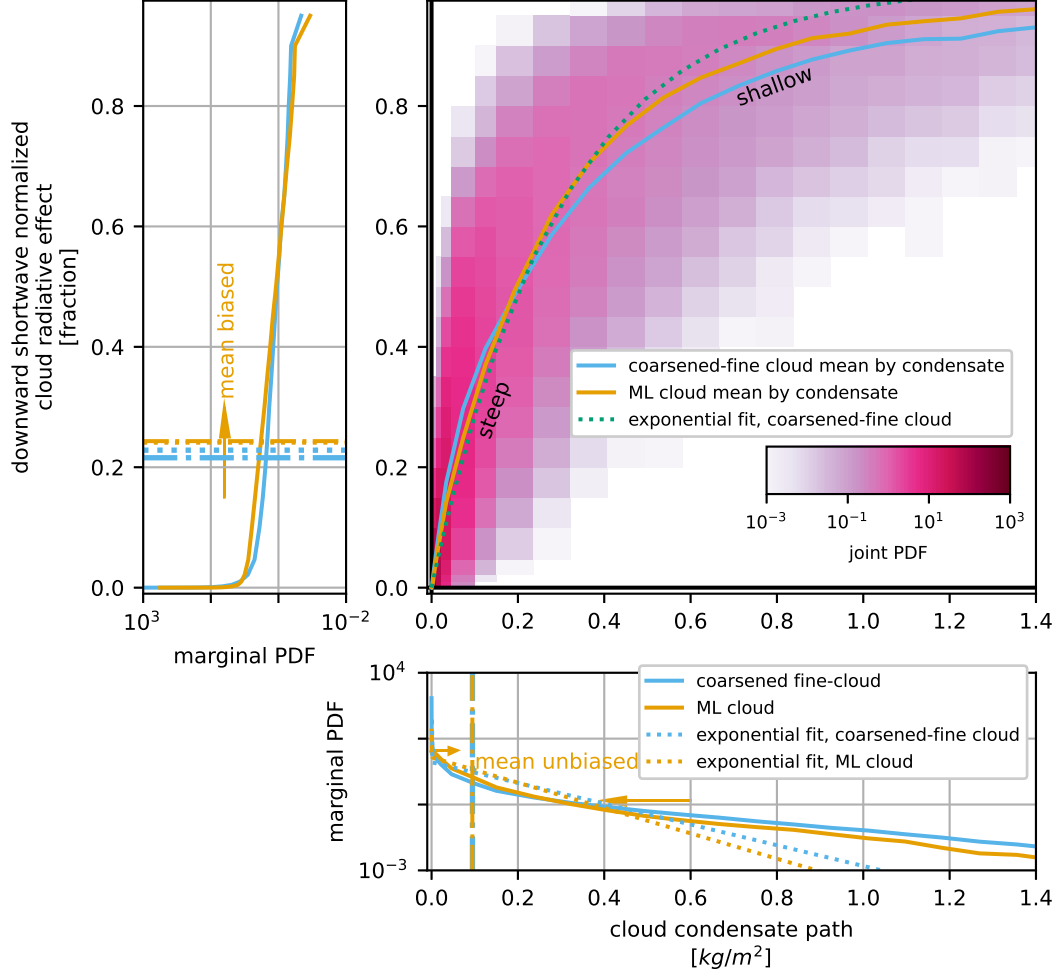
Fig. 9 shows the joint distribution of column total cloud condensate path (CCP) and  $\text{NSWCRE}^{sfc}$ . It has a convex shape, such that a given increase in CCP will increase  $\text{NSWCRE}^{sfc}$  much more for small CCP, i.e. predominantly thinner clouds. Thus, the mean  $\text{NSWCRE}^{sfc}$  of the ML clouds is larger than in the training data, because the ML cloud condensate shifts the CCP distribution towards its mean from both high and low values, but the shifts from low values have more impact on  $\text{NSWCRE}^{sfc}$ . In Appendix A, we show that simple exponential models of the distribution of CCP and its effect on  $\text{NSWCRE}^{sfc}$  can approximately predict the bias in  $\text{NSWCRE}^{sfc}$  resulting from an under-dispersed ML prediction of CCP.

### 3.3 Thresholded ML clouds

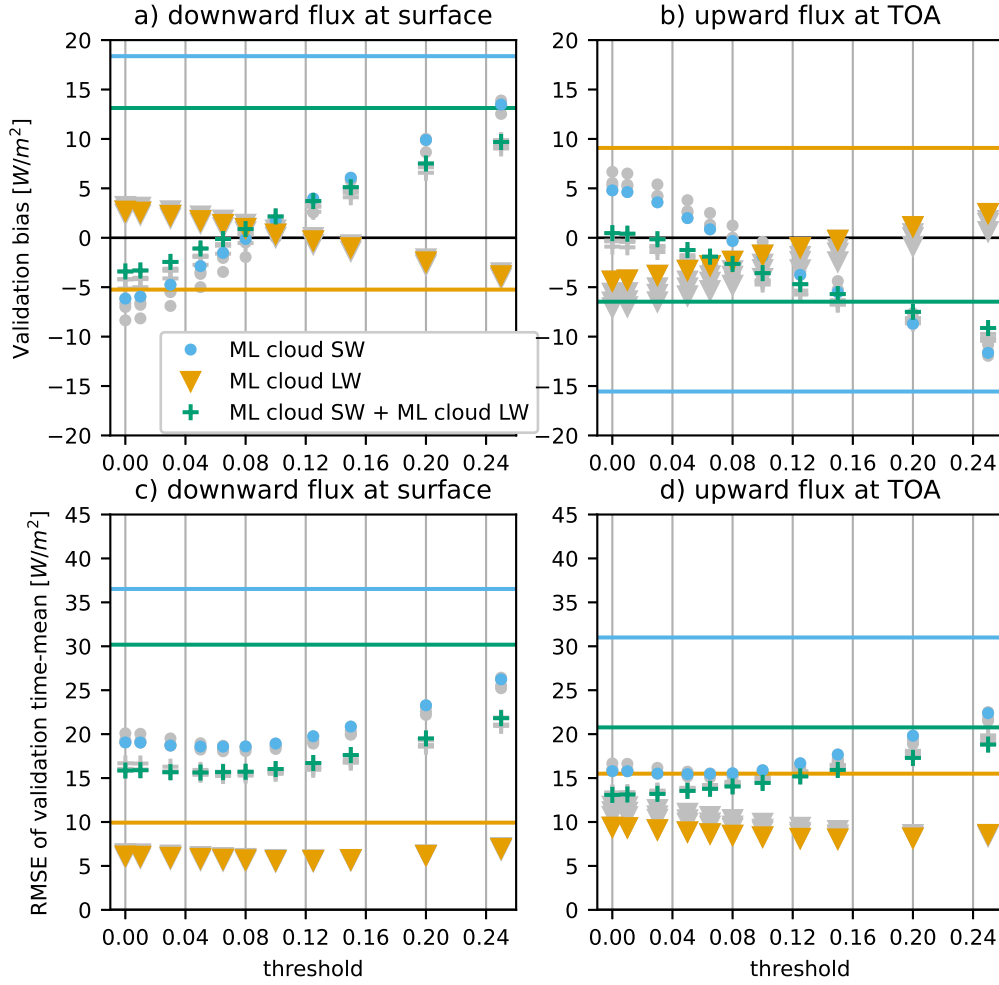
Due to the radiative flux biases resulting from imperfect distributions of ML cloud fields, we apply a post-processing approach, “thresholding” the raw ML cloud fields by setting to zero cloud condensate in all grid cells below a user-chosen threshold ML fractional cloud cover  $k$ . This mostly removes cloud from grid cells that have little condensate, which can have a meaningful radiative impact without inducing significant low biases in the global-mean condensate distribution. This simple approach can generate nearly unbiased and highly skillful radiative fluxes from our ML-predicted clouds.

Figs. 10a-b show the sensitivity of the global-time-mean TOA and surface radiative fluxes biases to the threshold  $k$ . The shortwave biases (blue circles) are larger than the partly compensating longwave biases (orange triangles). Their sum, the net flux, is shown as green pluses. Increasing  $k$  decreases the magnitude of both the shortwave and longwave biases, up to a point. Figs. 10c-d show the corresponding sensitivity of the RMSEs of the time-mean spatial pattern of the radiative fluxes. The RMSEs are fairly insensitive to  $k$ , but most show a slight thresholding-induced reduction in RMSE across a broad range of  $k$  including 0.065.

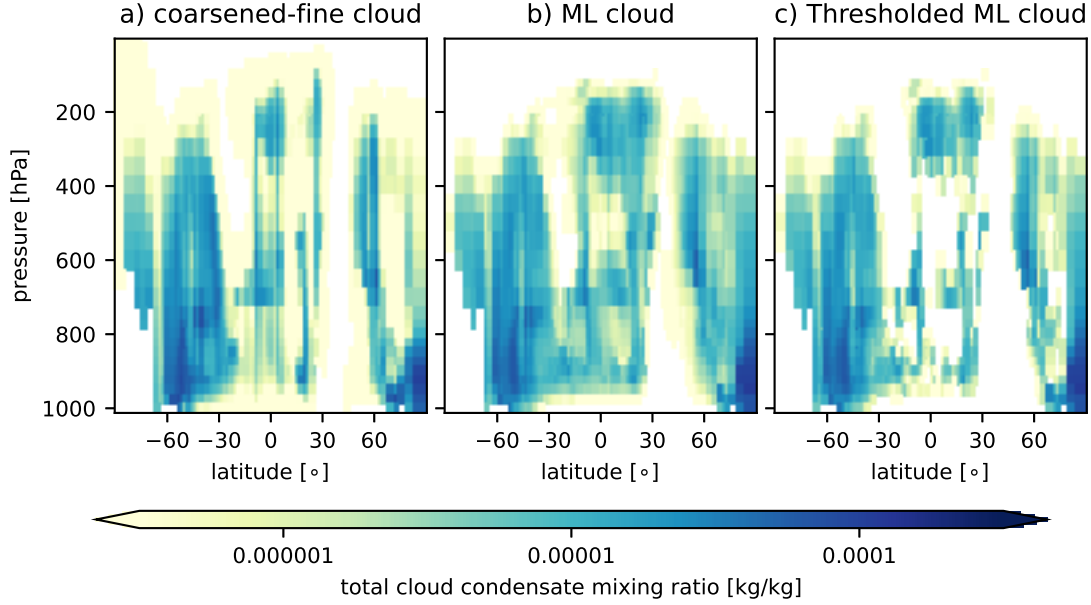
Using the selected seed model, we choose a threshold value ( $k = 0.065$ ) that most closely produces both unbiased net TOA radiation (shortwave plus longwave, an important goal in tuning climate models), and small magnitude shortwave and longwave component biases at both the surface and TOA. This choice results in net TOA bias of  $-2.8 \text{ W/m}^2$ , and component bias magnitudes all  $< 3 \text{ W/m}^2$ .



**Figure 9.** Main panel: joint distribution of CCP and NSWCRE<sup>sf</sup> in the coarsened-fine cloud training data. The mean of the training data and ML NSWCRE<sup>sf</sup> for each CCP bin are shown. The side panels show the marginal distributions of CCP and NSWCRE<sup>sf</sup> in the training data and ML. Orange arrows indicate shifts in the ML-predicted marginal PDFs of CCP and NSWCRE<sup>sf</sup> relative to the coarsened-fine PDFs. Exponential model best-fit lines are also shown with dotted lines (see Appendix A).



**Figure 10.** Effect of ML fractional cloud cover thresholding on radiative flux bias (a, b) and RMSE (c, d) at the surface (a, c) and TOA (b, d). Values are computed over the ML validation period. The NN random seed ensemble members are shown in gray, with the selected member plotted in color. The horizontal lines indicate the metrics from the nudged coarse baseline run.



**Figure 11.** Validation time-mean transects of target and ML cloud condensate for the north-south transect through the Indian Ocean shown in Fig. 2.

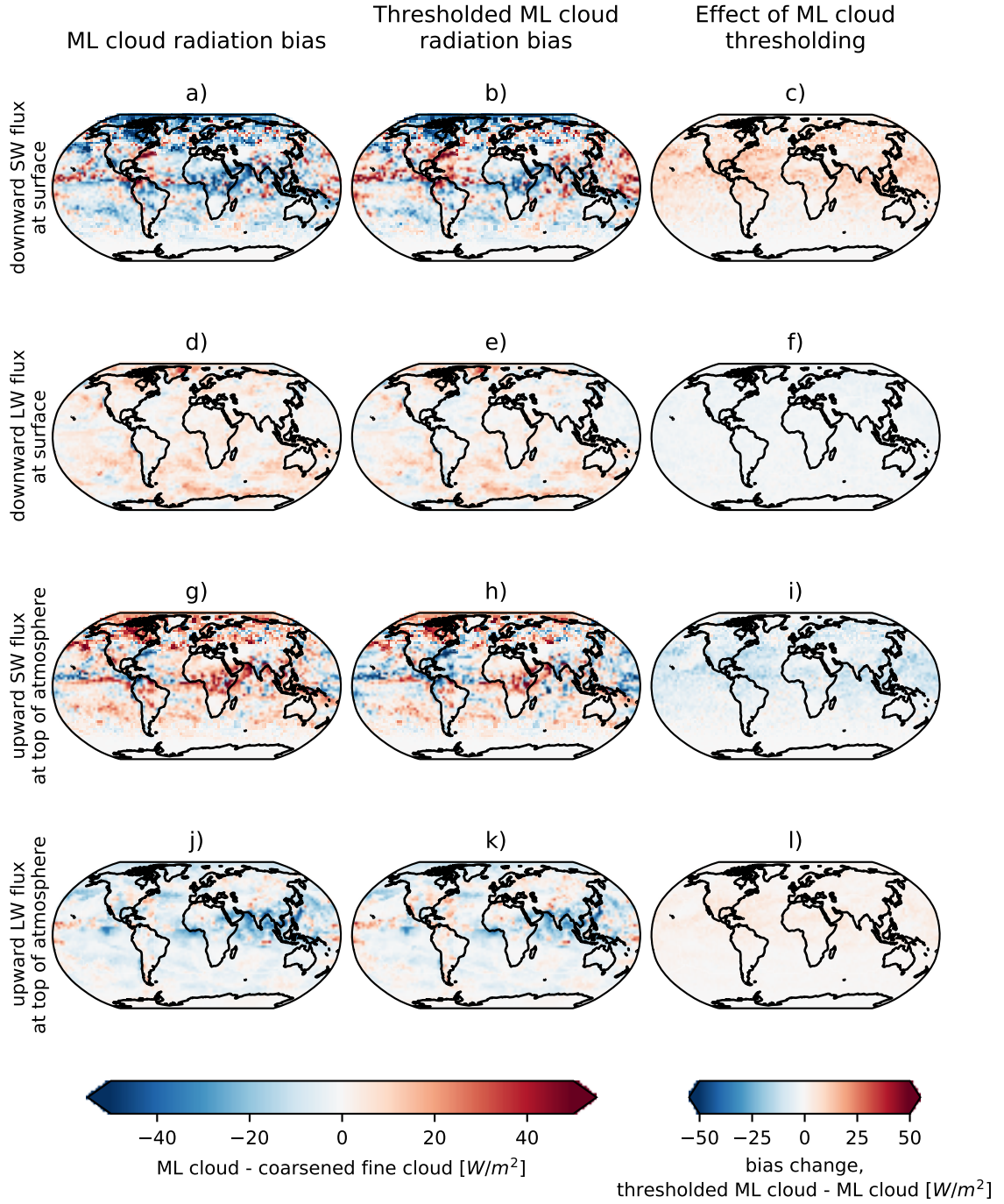
The effect of thresholding with  $k = 0.065$  on the global-mean vertical profiles and the probability distributions of ML cloud quantities was shown as orange dashed lines on Figs. 6 and 7. Thresholding introduces a negative bias of 5-10% into the post-processed ML cloud fields, but it has a negligible impact on their  $R^2$ . Significantly, thresholding produces a much better match to the fraction of clear columns in the training data (Fig. 7c). This translates into smaller radiative biases (last column, Table 2). Thus, thresholded ML cloud achieves the goal of producing nearly unbiased radiative fluxes from ML clouds, with much smaller error magnitudes than the baseline nudged coarse simulation.

Fig. 11 compares time-mean raw and thresholded ML cloud predictions against coarsened-fine validation data along the same vertical N-S transect shown in Fig. 2. The raw ML cloud predictions laterally spread condensate and contain too many slightly cloudy cells. The thresholded ML cloud improves the match to the validation data, although some regions of very thin, radiatively insignificant cloud with grid-mean condensate less than  $10^{-6}$  kg/kg in the validation transect are removed by the thresholding.

The radiation fields resulting from the raw and thresholded ML cloud are shown in Fig. 12. In the ML clouds' radiative fluxes, there is too little transmissivity in many columns, particularly those adjacent to areas of deep convection (Fig. 12a, g). The thresholding has the effect of reducing this bias, particularly for shortwave flux both at the surface and TOA (Fig. 12c, i). It does this without significantly worsening biases in areas where the raw ML cloud transmissivity is too high, i.e., in the specific columns with deep convection. It is less effective at reducing the longwave bias, particularly in upward flux at the TOA over the Indian Ocean and warm pool (Fig. 12j, k, l).

## 4 Discussion

While our ML cloud scheme outperforms its baseline and achieves nearly unbiased radiative fluxes, it is also deliberately simple. More sophisticated schemes might be able to predict cloud fields that produce unbiased radiative fluxes without a post-hoc thresh-



**Figure 12.** Global-mean radiative flux bias over the ML validation period. a, d, g, j: bias from ML clouds; b, e, h, k: bias from thresholded ML clouds; c, f, i, l: the change in bias due to thresholding the ML clouds at  $k = 0.065$ .



olding step. Several approaches might be employed to achieve this. For example, convolution of the model column via ML, rather than a gridcell-local approach, might result in better-predicted overlap structure of clouds and therefore not require removal of thin cloud in excess columns. Using ML to make an unbiased prediction of the joint PDF of fractional cloud cover and condensate in each grid cell (rather than the cloud fields themselves), and sampling appropriately from that PDF, might avoid the radiation biases associated with regression of ML-predicted condensate and cloud fraction toward their means. A related idea was described in Shamekh et al. (2022): during the coarsening of a fine-grid humidity field, a latent variable encoding subgrid organization was saved along with the grid-mean value; the latent variable added skill in predicting grid-mean precipitation rate. A similar latent variable could characterize subgrid cloud organization important for grid-mean radiative fluxes.

Radiative heating rates derived from ML cloud properties used prognostically (online) may improve coarse-grid simulations, given the baseline’s observed poor representation of cloud and radiation. While the corrective temperature tendency in nudging-based corrective ML already implicitly handles this (Bretherton et al., 2022), it would be more physically satisfying and consistent to attribute the difference in heating rates to a specific cloud bias. In the Supporting Information, we show that the vertically-resolved radiative shortwave heating rates from ML cloud are also more accurate than those from the coarse nudged baseline (Fig. S4), suggesting that vertical cloud placement is improved over the baseline. However, this is not necessarily the case for ML cloud in terms of long-wave heating rates (Fig. S5).

One of the inputs to the cloud ML model is the nudged coarse model’s ice cloud condensate mixing ratio. Unlike the other ML inputs (temperature, pressure, and relative humidity) that are coarse model thermodynamic state variables, the ice cloud condensate comes from its physics parameterization, and is included because it may capture non-local effects on cloud such as convective updrafts. While including this feature does improve the ML predictions of coarsened-fine ice cloud condensate and upward long-wave fluxes at TOA, the improvement is marginal and not a requirement for the overall goal of improving coarse model radiative fluxes. This is helpful as the behavior of coarse model physics may not be robust across models and configurations.

## 5 Conclusions

Coarse-grid weather and climate models rely on parameterizations of the subgrid variability of cloud fields, and coarse-grained cloud fields from a fine-grid storm-resolving reference model are a natural target for a data-driven (ML) parameterization. We implement this approach in a 200 km grid global atmospheric model, FV3GFS, with ML trained on coarsened outputs (grid-mean cloud fraction, liquid and ice condensate) from a reference global 3-km grid simulation using a modified version of FV3GFS, X-SHIELD. These outputs are used in the FV3GFS radiation scheme, and our goal was to obtain accurate radiative fluxes and heating rates from the learned cloud properties.

With an appropriate vertical overlap scheme and a physically correct setting of a GFS physics parameter called “ccnorm” (which is set incorrectly in NOAA’s current operational forecast versions of this model), the coarsened fine-grid clouds from a fine-grid reference model produce almost unbiased surface and TOA radiative fluxes when used in the coarse-grid radiation parameterization.

The ML skillfully learns the coarsened-fine cloud properties as a function of local coarse-grid model state. But because the ML is not perfect and the radiative effects of a cloud layer depend non-linearly on its thickness, the global-mean TOA radiative fluxes derived from the machine-learned clouds are biased, even though the predicted cloud fields themselves are not. We show that zeroing predicted cloud condensate in cells with an

ML-predicted cloud fraction less than a threshold of 0.065 largely removes these biases with minimal impacts to the skill of the cloud predictions of the ML scheme. The resulting ML-derived radiative fluxes are much more accurate than those produced by the existing cloud parameterization in the nudged coarse-grid model.

To be an attractive candidate for on-line implementation in the coarse model, an ML scheme for clouds would also have to produce physically justifiable precipitation fields that approximately match the coarsened fine-grid reference data. This would be an excellent extension of our work.

## Acknowledgments

We acknowledge the Allen Institute for Artificial Intelligence (AI2) for support of this project. We acknowledge the NOAA Geophysical Fluid Dynamics Laboratory (GFDL) for providing the computational resources to perform the fine-grid reference simulations. We also acknowledge the scientific and software support provided by AI2 climate modeling team members.

## Open Research

A snapshot of the Python and Fortran codes used for the ML and diagnostic radiation workflows, running FV3GFS, and making figures and tables is preserved at <https://zenodo.org/record/8176388> via MIT License (Henn et al., 2023). The code is also available on GitHub at <https://github.com/ai2cm/radiation-cloud-ML-workflow>. The fine-grid model restart and diagnostic files spanning 10 days at 15 minute intervals, coarsened to C48 resolution (about 120 GB), are needed for the nudged and prescribed-cloud FV3GFS runs. They are available on Google Cloud Storage upon request to the corresponding author, via a publicly available “requester pays” bucket.

## References

- Brenowitz, N. D., & Bretherton, C. S. (2019, 8). Spatially extended tests of a neural network parametrization trained by coarse-graining. *Journal of Advances in Modeling Earth Systems*, 11, 2728-2744. Retrieved from <https://onlinelibrary.wiley.com/doi/10.1029/2019MS001711> doi: 10.1029/2019MS001711
- Bretherton, C. S., Henn, B., Kwa, A., Brenowitz, N. D., Watt-Meyer, O., McGibbon, J., ... Harris, L. (2022, 2). Correcting coarse-grid weather and climate models by machine learning from global storm-resolving simulations. *Journal of Advances in Modeling Earth Systems*, 14. Retrieved from <https://onlinelibrary.wiley.com/doi/10.1029/2021MS002794> doi: 10.1029/2021MS002794
- Caldwell, P. M., Zelinka, M. D., Taylor, K. E., & Marvel, K. (2016). Quantifying the sources of intermodel spread in equilibrium climate sensitivity. *Journal of Climate*, 29, 513-524. doi: 10.1175/JCLI-D-15-0352.1
- Chantry, M., Hatfield, S., Dueben, P., Polichtchouk, I., & Palmer, T. (2021, 7). Machine learning emulation of gravity wave drag in numerical weather forecasting. *Journal of Advances in Modeling Earth Systems*, 13. doi: 10.1029/2021MS002477
- Chen, G., Wang, W.-C., Yang, S., Wang, Y., Zhang, F., & Wu, K. (2023, 4). A neural network-based scale-adaptive cloud-fraction scheme for gcms. Retrieved from <http://arxiv.org/abs/2304.01879>
- Cheng, K. Y., Harris, L., Bretherton, C., Merlis, T. M., Bolot, M., Zhou, L., ... Fueglistaler, S. (2022, 8). Impact of warmer sea surface temperature on the global pattern of intense convection: Insights from a global storm resolving

- model. *Geophysical Research Letters*, 49. doi: 10.1029/2022GL099796
- Clark, S. K., Brenowitz, N. D., Henn, B., Kwa, A., McGibbon, J., Perkins, W. A.,  
... Harris, L. M. (2022). Correcting a 200 km resolution climate model in multiple climates by machine learning from 25 km resolution simulations. *Journal of Advances in Modeling Earth Systems*, 14. doi: 10.1029/2022MS003219
- Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., & Yacalis, G. (2018). Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, 45, 5742-5751. doi: 10.1029/2018GL078202
- Grundner, A., Beucler, T., Gentine, P., & Eyring, V. (2023, 4). Data-driven equation discovery of a cloud cover parameterization. Retrieved from <http://arxiv.org/abs/2304.08063>
- Grundner, A., Beucler, T., Gentine, P., Iglesias-Suarez, F., Giorgetta, M. A., & Eyring, V. (2022, 12). Deep learning based cloud cover parameterization for icon. *Journal of Advances in Modeling Earth Systems*, 14. doi: 10.1029/2021MS002959
- Harris, L., Zhou, L., Lin, S. J., Chen, J. H., Chen, X., Gao, K., ... Stern, W. (2020). Gfdl shield: A unified system for weather-to-seasonal prediction. *Journal of Advances in Modeling Earth Systems*, 12. doi: 10.1029/2020MS002223
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference and prediction* (Vol. 2nd ed.). Springer-Verlag.
- Henn, B., Jauregui, Y. R., Clark, S. K., Brenowitz, N. D., McGibbon, J., Watt-Meyer, O., ... Bretherton, C. S. (2023, July). *Code for "A machine learning parameterization of clouds in a coarse-resolution climate model for unbiased radiation", a manuscript submitted to JAMES*. Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.8176388> doi: 10.5281/zenodo.8176388
- Hu, Y. X., & Stamnes, K. (1993, 4). An accurate parameterization of the radiative properties of water clouds suitable for use in climate models. *Journal of Climate*, 6, 728-742. Retrieved from [http://journals.ametsoc.org/doi/10.1175/1520-0442\(1993\)006<0728:AAPOTR>2.0.CO;2](http://journals.ametsoc.org/doi/10.1175/1520-0442(1993)006<0728:AAPOTR>2.0.CO;2) doi: 10.1175/1520-0442(1993)006<0728:AAPOTR>2.0.CO;2
- Keller, A. C., & Evans, J. M. (2019). Application of random forest regression to the calculation of gas-phase chemistry within the geos-chem chemistry model v10. *Geoscientific Model Development*, 12, 1209-1225. doi: 10.5194/gmd-12-1209-2019
- Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization*. arXiv. Retrieved from <https://arxiv.org/abs/1412.6980> doi: 10.48550/ARXIV.1412.6980
- Krasnopolsky, V. M., Fox-Rabinovitz, M. S., & Chalikov, D. V. (2005). *New approach to calculation of atmospheric model physics: Accurate and fast neural network emulation of longwave radiation in a climate model*.
- Krasnopolsky, V. M., Fox-Rabinovitz, M. S., Hou, Y. T., Lord, S. J., & Belochitski, A. A. (2010, 5). Accurate and fast neural network emulations of model radiation for the ncep coupled climate forecast system: Climate simulations and seasonal predictions. *Monthly Weather Review*, 138, 1822-1842. doi: 10.1175/2009MWR3149.1
- Kwa, A., Clark, S. K., Henn, B., Brenowitz, N. D., McGibbon, J., Watt-Meyer, O., ... Bretherton, C. S. (2023, 5). Machine-learned climate model corrections from a global storm-resolving model: Performance across the annual cycle. *Journal of Advances in Modeling Earth Systems*, 15. Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003400> doi: 10.1029/2022MS003400
- Lagerquist, R., Turner, D., Ebert-Uphoff, I., Stewart, J., & Hagerty, V. (2021, 7). Using deep learning to emulate and accelerate a radiative-transfer model. *Journal of Atmospheric and Oceanic Technology*. Retrieved from <https://journals.ametsoc.org/view/journals/atot/aop/JTECH-D-21-0007.1/>

- JTECH-D-21-0007.1.xml doi: 10.1175/JTECH-D-21-0007.1
- Liu, Q., & Yang, F. (2023). Solar and thermal infrared radiation transfer schemes in the ncep global forecast system (gfs). Retrieved from <https://doi.org/10.25923/cryn-tp50> doi: 10.25923/cryn-tp50
- McGibbon, J., & Bretherton, C. S. (2019). Single-column emulation of reanalysis of the northeast pacific marine boundary layer. *Geophysical Research Letters*, 46(16), 10053-10060. Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL083646> doi: <https://doi.org/10.1029/2019GL083646>
- Mlawer, E. J., Iacono, M. J., Pincus, R., Barker, H. W., Oreopoulos, L., & Mitchell, D. L. (2016, 4). Contributions of the arm program to radiative transfer modeling for climate and weather applications. *Meteorological Monographs*, 57, 15.1-15.19. doi: 10.1175/amsmonographs-d-15-0041.1
- O’Gorman, P. A., & Dwyer, J. G. (2018, 10). Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, 10, 2548-2563. doi: 10.1029/2018ms001351
- Perkins, W. A., Brenowitz, N. D., Bretherton, C. S., & Nugent, J. M. (2023). Emulation of cloud microphysics in a climate model. doi: 10.22541/essoar.168614667.71811888/v1
- Pincus, R., Barker, H. W., & Morcrette, J. J. (2003). A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields. *Journal of Geophysical Research: Atmospheres*, 108. doi: 10.1029/2002jd003322
- Putman, W. M., & Lin, S.-J. (2007). Finite-volume transport on various cubed-sphere grids. *Journal of Computational Physics*, 227(1), 55 - 78. doi: <https://doi.org/10.1016/j.jcp.2007.07.022>
- Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. Retrieved from <https://gitlab.com/mspritch/spcam3.0-neural-net/tree/nn> doi: 10.5281/zenodo.1402384
- Shamekh, S., Lamb, K. D., Huang, Y., & Gentine, P. (2022). Implicit learning of convective organization explains precipitation stochasticity. doi: 10.1002/essoar.10512517.1
- Watt-Meyer, O., Brenowitz, N. D., Clark, S. K., Henn, B., Kwa, A., McGibbon, J., ... Bretherton, C. S. (2021, 8). Correcting weather and climate models by machine learning nudged historical simulations. *Geophysical Research Letters*, 48, 1-13. Retrieved from <https://onlinelibrary.wiley.com/doi/10.1029/2021GL092555> doi: 10.1029/2021GL092555
- Yuval, J., O’Gorman, P. A., & Hill, C. N. (2021). Use of neural networks for stable, accurate and physically consistent parameterization of subgrid atmospheric processes with good performance at reduced precision. *Geophysical Research Letters*, 48, 1-11. doi: 10.1029/2020GL091363
- Zhou, L., Harris, L., Chen, J. H., Gao, K., Guo, H., Xiang, B., ... Morin, M. (2022, 7). Improving global weather prediction in gfdl shield through an upgraded gfdl cloud microphysics scheme. *Journal of Advances in Modeling Earth Systems*, 14. doi: 10.1029/2021MS002971
- Zhou, L., Lin, S. J., Chen, J. H., Harris, L. M., Chen, X., & Rees, S. L. (2019, 7). Toward convective-scale prediction within the next generation global prediction system. *Bulletin of the American Meteorological Society*, 100, 1225-1243. doi: 10.1175/BAMS-D-17-0246.1

## Appendix A A simple exponential model for the effects of cloud field distribution on radiative flux

The bottom panel of Fig. 9 suggests that the PDF of column total cloud condensate path (CCP) can be fitted by a truncated exponential distribution of the form:

$$p(CCP) = \frac{c_m}{c_1^2} e^{-CCP/c_1}, CCP > 0, c_m < c_1. \quad (A1)$$

The first fitting parameter is the characteristic CCP decay scale  $c_1$ . The second fitting parameter  $c_m$  is the mean of the CCP PDF. This PDF integrates over  $CCP > 0$  to the fraction of cloud-containing columns  $c_m/c_1 < 1$ . The remainder of the PDF, which falls at  $CCP = 0$ , and represents the fraction of clear columns.

The main panel of Fig. 9 suggests that the normalized cloud effect on surface downwelling shortwave radiation,  $NSWCRE^{sf c}$  can also be fitted as an exponential function of CCP:

$$NSWCRE^{sf c}(CCP) = 1 - e^{-CCP/c_0}, \quad (A2)$$

where the fitting parameter  $c_0$  is a characteristic value of CCP.

Combining these two empirical results gives a simple mathematical model that predicts  $NSWCRE^{sf c}$ . This enables us to predict the extent to which an ML-predicted distribution of CCP, which is unbiased in the mean (correct  $c_m$ ) but inherently under-dispersed due to imperfect ML skill (underestimated  $c_1$ ), would overestimate  $NSWCRE^{sf c}$ .

Combining eqs. A2 and A1 with the expectation operator over CCP, we get:

$$\begin{aligned} E[NSWCRE^{sf c}] &= \int_0^\infty NSWCRE^{sf c}(CCP) \cdot p(CCP) \cdot dCCP \\ &= \int_0^\infty (1 - e^{-CCP/c_0}) \cdot \frac{c_m}{c_1^2} e^{-CCP/c_1} \cdot dCCP \\ &= \frac{c_m}{c_0 + c_1} \end{aligned} \quad (A3)$$

For given values of the mean column condensate path  $c_m$  and the radiative fit parameter  $c_0$ , the global-mean  $NSWCRE^{sf c}$  is a decreasing function of  $c_1$ . An ML fit which reduces  $c_1$  will inevitably increase  $NSWCRE^{sf c}$ . This is exactly the effect we wished to isolate.

For our coarsened-fine cloud dataset, the global-mean CCP is  $c_m = 0.095 \text{ kg/m}^2$ , and based on the data in Fig. 9, the radiative fit parameter  $c_0 = 0.30 \text{ kg/m}^2$ . We calibrate  $c_m/c_1$  to match the fraction of coarse grid columns with radiatively significant  $CCP > 1e^{-3} \text{ kg/m}^2$ ; this yields  $c_1^{ref} = 0.118 \text{ kg/m}^2$  (for which 80% of coarse grid columns have radiatively significant cloud), and  $c_1^{ML} = 0.097 \text{ kg/m}^2$  (for which 98% of coarse grid columns have radiatively significant cloud).

Finally, applying eq. A3 using these values yields expected  $NSWCRE^{sf c}$  of 0.241 for ML vs. 0.228 for the coarsened-fine data, i.e., these exponential models predict a 5.4% overestimation of  $NSWCRE^{sf c}$  due to the under-dispersion of the ML predictions (Fig. 9, dashed lines on left panel). In fact, the actual overestimation of  $NSWCRE^{sf c}$  is 12.8% for the ML versus coarsened-fine datasets (0.243 vs 0.216, Fig. 9, solid lines on left panel). The expected bias will also depend on the altered exponential relationship between CCP and  $NSWCRE^{sf c}$  in the predicted cloud due to differences in vertical overlap, which we neglect here to focus on the effects of under-dispersion and which has a relatively small on global-mean  $NSWCRE^{sf c}$ . However, these numbers show that the exponential model fits the data quite well.