Scale-aware parameterization of cloud fraction and condensate for a global atmospheric model machine-learnt from coarse-grained kilometer-scale simulations

Cyril Morcrette[1,2,3,4], Tobias Cave[5], Helena Reid[1], Joana da Silva Rodrigues[1], Teo Deveney[5], Lisa Kreusser[5], Kwinten Van Weverberg[1,6,7], Chris Budd[5]

[1]Atmospheric Processes and Parameterizations, Met Office, Exeter, UK
[2]Department of Mathematics and Statistics, University of Exeter, Exeter, UK
[3]Global Systems Institute, University of Exeter, Exeter, UK
[4]Institute for Data Science and Artificial Intelligence, University of Exeter, Exeter, UK
[5]Department of Mathematical Sciences, University of Bath, Bath, UK
[6]Royal Meteorological Institute of Belgium, Brussels, Belgium
[7]Department of Geography, Ghent University, Ghent, Belgium

Key Points:

- High-resolution weather simulations used as training data for climate model development.
- Neural network trained on coarse-grained data predicts cloud cover, liquid water content, ice water content all as a function of height.
- Runs stably when coupled interactively to a climate model.

Corresponding author: Cyril Morcrette, `cyril.morcrette@metoffice.gov.uk`

Abstract

Kilometer grid-length simulations over a variety of different locations worldwide are used as training data for a deep-learning model designed to predict clouds in a global climate model. The inputs to the neural network are profiles of temperature, humidity and pressure from the high-resolution model, averaged to the scale of the climate model. The outputs are profiles of cloud fraction, liquid water content and ice water content. The high-resolution data is coarse-grained to a range of sizes, allowing the model to learn how the cloud formation depends on the size of the area being considered. The machine-learnt cloud cover and condensate scheme is coupled to a global climate model and used to run multi-year simulations where the clouds predicted by the neural-network are fully interacting with the rest of the model.

Plain Language Summary

Clouds play an important part in our climate. Their effects on sunshine, trapped heat and rainfall need to be represented accurately in the computer models used to make predictions about our future climate. However most climate models use a grid spacing that is much coarser than the size of clouds. The traditional way of dealing with this is to write mathematical descriptions of how clouds form based on temperature and humidity and include them in the climate model. As an alternative, we run a series of much higher detail weather-forecast-like simulations. These are known to be better at predicting cloudiness, and then we average the data to the scale of a climate model. Some machine learning techniques are then used to predict the cloud cover and the amount of liquid and ice they contain. This machine-learnt method is then put into a climate model where the clouds it produces interacts with sunlight and lead to rainfall production. The hybrid model is able to run effectively for several years.

1 Introduction

Clouds are an important part of the Earth's climate and improving their representation in climate models is crucial for reducing the uncertainty in climate projections (Bony et al., 2015).

The physical processes that lead to cloud formation occurs on scales that can vary widely. In frontal zones, the large-scale ascent and cloud formation can occur over hundreds of kilometers and global climate models with grid-lengths of 50 to 150 km can represent them relatively well. There are many cases however where cloud formation is determined by processes occurring on scales much smaller than the model grid-spacing. If one does nothing about this separation between the scale of the numerical model and the scale of the cloud formation process, and only allows a model to represent clouds when saturation occurs on their resolved grid, then the simulations would be very poor. Indeed, they would fail to represent the effects that clouds have on surface temperature and precipitation and the feedbacks that they exert on the climate system (Bony et al., 2015).

As a result, large-scale cloud-cover parameterizations have been developed to ensure that the radiative and microphysical impact of clouds are included in climate simulations. Additionally, convection schemes have been created to represent the removal of atmospheric instabilities and the vertical redistribution of mass, humidity and momentum. These schemes also estimate convective cloudiness, which contributes to the atmospheric radiative balance.

Stensrud (2007) provides a general review of parameterization schemes used in weather and climate models. For cloud cover specifically, some of the earliest approaches are described by Sommeria and Deardorff (1977) and Sundqvist (1978). Even though a grid

box may be sub-saturated in a gridbox mean sense, variability in temperature and humidity may lead to local saturation and cloud formation within some part of the grid box. This leads to the concept of critical relative humidity, the threshold, below 100%, at which clouds are allowed to form in a model. The representation, and evaluation, of clouds in atmospheric models has received a lot of attention (Slingo, 1987; Randall, 1989; Jakob & Klein, 1999; Randall et al., 2003; Illingworth et al., 2007; Morcrette et al., 2012). Progressively more complicated and accurate schemes have been developed (Smith (1990); Le Treut and Li (1991); Tiedtke (1993); Xu and Randall (1996); Golaz et al. (2002); Tompkins (2002); Wilson, Bushell, Kerr-Munslow, Price, and Morcrette (2008); Watanabe et al. (2009); Van Weverberg et al. (2016, 2021)).

Meanwhile, the field of machine learning and its application to atmospheric modeling has been rapidly evolving. One area is the impressive development of emulators that can be used to replace the entirety of a numerical weather prediction model. Over the last few years, these methods have caught up and surpassed the performance of traditional approaches for global weather forecasting (Dueben & Bauer, 2018; Weyn et al., 2020; Keisler, 2022; Bi et al., 2023; Lam et al., 2023; Husain et al., 2024). Most recently, Kochkov et al. (2024) replaced a model's physics parameterizations with machine-learnt algorithms, producing ensemble forecasts that can outperform both traditional physical model ensembles and deterministic emulators. They also showed that a machine learning model trained on short timescale data can remain stable over climate timescales.

A comprehensive review of intersection between machine learning and weather and climate modelling is provided by de Burgh-Day and Leeuwenburg (2023). Machine learning can be used to develop improvements to individual parameterization schemes. This includes the work of Chevallier et al. (1998) and Krasnopolsky et al. (2005) in the field of long-wave radiative transfer, Krasnopolsky et al. (2013), Ukkonen and Mäkelä (2019) and Kumar et al. (2024) for convective parameterizations, and Seifert and Rasp (2020) and Chantry et al. (2021) for microphysics and gravity-wave drag schemes respectively.

Coarse-resolution models are very common in weather and climate modelling due to their computational affordability. One way to improve coarse-resolution models is to use super-parameterization. This consists in running a high-resolution model within each coarse model gridbox and using coarse-grained high-resolution tendencies to update the coarse-resolution model (Khairoutdinov et al., 2005). Due to the computational cost, the high-resolution model is typically run with certain simplifications, such as being two-dimensional or having a width narrower than the coarse model gridbox. Such an approach can improve the simulations of precipitation and reduce biases seen in the coarse model (Kooperman et al., 2016).

The use of high-resolution simulations and machine learning can be brought together to form a new approach to model development (Schneider et al., 2017). Gentine et al. (2018) showed that inputs and outputs of high-resolution super-parameterizations could be used as a training data to enable machine-learning to emulate the behavior of the embedded high-resolution model. When the high-resolution emulator was incorporated into the coarse model, Rasp et al. (2018) showed that the benefits of the super-parameterizations were retained but without the increased computational runnning cost. Similar conclusions were drawn by Brenowitz and Bretherton (2018, 2019) who ran kilometer-scale simulations, machine-learnt coarse-grained high-resolution thermodynamic tendencies and embedded them in a global model, leading to improvements in some biases compared to the control coarse model.

This study is focused on using high-resolution simulations of clouds, to machine-learn a new parameterization of clouds for use in a climate model as a route to improving climate projections. Building on the work of Lewis et al. (2022), output from the Met Office Unified Model (MetUM) is used as training data for developing a machine-learnt algorithm that predicts clouds. The approach is similar to that taken by Grundner et

al. (2022), who used kilometer-scale data to build an emulator that could predict the fractional cloud cover through a model column. Our approach is different in a few aspects. Firstly, we simultaneously emulate the liquid and ice water content, as well as the fractional cloudiness ensuring that information about both cloud coverage and condensate amounts from the high-resolution simulations is incorporated into the climate model. Secondly, in contrast to previous works our high-resolution model data is coarse-grained to a range of coarser resolutions and a measure of the coarse-graining size is used as an input during training. As a result, the emulator has the potential to learn scale-aware behavior allowing it to be deployed in models with a non-uniform mesh, such as a latitude-longitude grid. Finally, a key development in this work, is the interactive deployment of the machine-learnt cloud scheme in a climate model to replace the previous scheme which calculated cloud cover and condensate amounts. Our machine-learnt cloud cover and liquid and ice water contents interact with the radiation and microphysics schemes, release latent heat and the whole simulation is able to run stably for many years.

The rest of this paper is organized as follows. Section 2 describes the methods for generating and pre-processing the data used for training and outlines the architecture used in the emulator. Results are presented in section 3 and discussed in section 4. Finally, conclusions are drawn in section 5.

2 Methods

In this section, we describe the generation and pre-processing of the training data, as well as the design of the neural network architecture.

2.1 Generation of training data using MetUM

The data used for training are generated using the MetUM, which can be run as a "nesting suite" (Webster et al., 2008), where global model hindcasts are run from analyses to produce initial and lateral boundary conditions for some limited-area models (LAMs). The global driving model employs the Global Atmosphere 6 configuration (Walters et al., 2017, GA6), which has been used for operational global forecasting. GA6 has parameterizations for short and long-wave radiative heating, surface exchanges, boundary-layer mixing, flow over orography, convection, and cloud and precipitation formation. The global model is run using a resolution known as N1280, which consists of a regular latitude-longitude grid with 2560 points of longitude and 1920 points of latitude. This leads to a longitude spacing of 0.14° and an east-west grid spacing of 15 km on the equator (Walters et al., 2017), with progressively finer grid boxes nearer the poles. The global driving model has a timestep of 6 minutes and updates the lateral boundary conditions once per hour.

The LAMs use a grid-length of 1.5 km and the parameterizations used therein are described by Bush et al. (2023). The key difference between the LAMs and the driving model is that the LAMs are run without a convection scheme. This LAM configuration is known to perform better than the coarser model using a convective parameterization, when simulating convection over the United Kingdom (Lean et al., 2008) or over tropical western Africa (Stein et al., 2015).

The global version of the MetUM driving a regional configuration has been used to study a range of atmospheric processes on the kilometer-scale. This includes severe weather over New Zealand (Webster et al., 2008), strong winds over Antarctica (Orr et al., 2013), convective systems over the American Midwest (Kain et al., 2017) and South Africa (Keat et al., 2019), precipitation over South Korea (Song et al., 2019) and the interactions between clouds and radiation over polar, tropical and extra-tropical locations (Van Weverberg & Morcrette, 2022). As a result, the suite is known to perform well when relocated to various regions around the world. Typically the nesting suite is used to run a few LAMs in a handful of different locations. This nesting is one-way, i.e. the high-
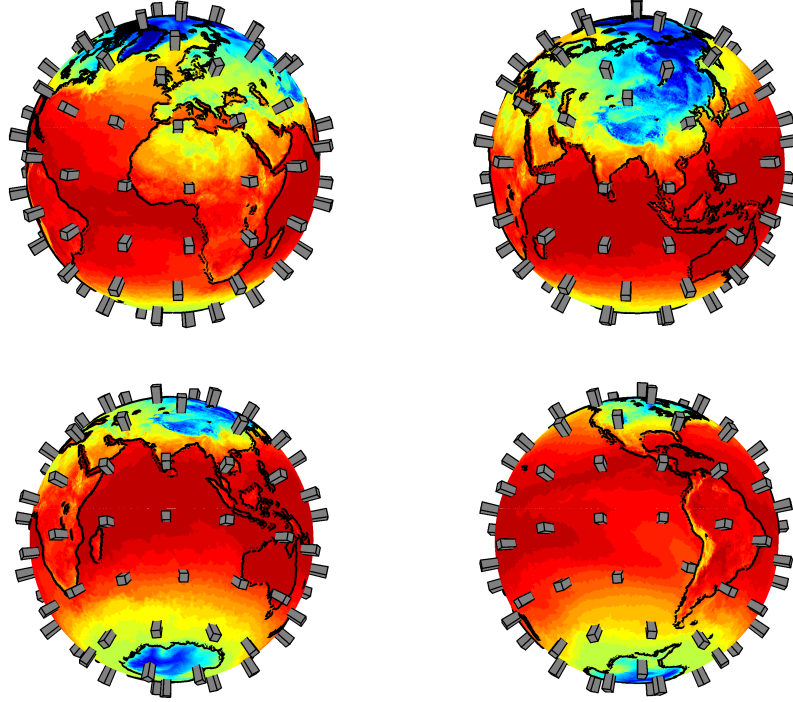
Figure 1. Location of the 80 limited-area models (LAMs) across the globe. Each LAM is 768 x 768 km wide (512 x 512 grid-points with a horizontal gridlength of 1.5 km). The shading is the near-surface temperature at 00Z on 1 Jan 2020. The height of each LAM has been exaggerated by a factor of 10 for display purposes.

resolution model does not feed back on the coarser model. In this study, the nesting suite is extended to deal with 80 separate LAMs.

Each LAM is defined to be 512×512 grid-points in the horizontal. Although using a latitude-longitude grid, each LAM uses a rotated pole such that the horizontal gridlength is an almost uniform 1.5 km. The LAMs use a time-step of 1 minute and all diagnostics are output every 2 hours. The locations of the 80 nested LAMs are shown by the black polygons in Fig. 1. These were chosen by projecting the globe's surface into an 20-faced polygon (icosahedron) and then splitting each equilaterial triangular face into 4, placing a LAM at the centre of each. For 4 of the 80 domains, the initial location had to be moved by roughly one degree to avoid steep orography or the inclusion of islands that were too small to be present in the driving model and thus made it difficult to generate surface properties with the nesting suite.

The MetUM uses a stretched vertical grid, with finer resolution lower in the atmosphere. The global driving model uses 70 stretched vertical levels, distributed with 50 levels below 18 km, and a model top at 80 km (Walters et al., 2017). Usually, the regional version of the MetUM is run with a lower lid at 40 km, so that the 70 levels are compressed, providing extra resolution in the boundary layer and lower troposphere at the expense of resolving stratospheric processes. Here, however, the same vertical level set is used in the LAMs as in the driving model to ensure that the machine-learning algorithm is trained on data of comparable vertical resolution to that of the global model it would be implemented back into.

Although the simulations for generating the training data are being performed at coarser vertical resolution than what is normal, it means that the driving model and LAM data can be compared without interpolation. Any machine-learning model can then be implemented back into the atmospheric model without dealing with the extra complexity of different level sets.

It is worth noting that all LAMs use the same first Regional Atmosphere and Land configuration (Bush et al., 2023, RAL2), specifically the tropical configuration (RAL2-T). The advantage of using RAL2-T is that it uses the same prognostic large-scale cloud parameterization scheme as the global model (Wilson, Bushell, Kerr-Munslow, Price, & Morcrette, 2008), unlike RAL2-M which uses a diagnostic approach.

Each of the 80 LAMs is free-running for the month of January 2020, with no re-initialisation within the domains. The global driving model, which provides the lateral boundary forcing, is re-initialised every 24 hours using 00Z operational MetUM atmospheric analyses. Daily sea surface temperatures are provided by the OSTIA system (Donlon et al., 2012).

2.2 Pre-processing of training data

It takes a certain amount of time, and distance, for fine-scale details to spin-up within each of the LAMs. To account for temporal spin-up, the simulation actually starts on the last day of December and that first day is discarded due to spin-up. For spatial spin-up due to coarse-scale information from the global model being fed in from the lateral boundaries, inspection of various model fields indicates that this occurs within a distance of around 10 gridpoints from the outer rim. As a result, a 16-gridpoint-wide region across all 4 edges of each LAM is ignored. The remaining 480 x 480 grid-point portion of each LAM is coarse-grained to a range of scales (Fig 2).

In Fig. 2b) this is done by averaging over 96 x 96 gridbox regions. With a model grid-length of 1.5 km, this corresponds to averaging over 144 km, comparable to a typical global climate model running with a 1 degree grid. A total of 25 such averaging regions can be found within the 480 x 480 central portion of each LAM. Progressively smaller regions are then considered, each with 64, 48, 32, 24, 18, 12, 9 and 6 gridboxes on each side, some of which are shown in Fig. 2. Although more averaging regions of smaller sizes can be fitted into each LAM, this is intentionally not done in order to avoid generating an unbalanced data set. Indeed, generating as many averages as possible would lead to 256 times more samples valid for a scale of 9km than for a scale of 144km. This could easily lead to a network that is biased towards predicting clouds better at small scales. Although this could be dealt with by a judicious choice of cost function, it is more straightforward to design the coarse-graining to be balanced in the first place. The last step in pre-processing is to rescale all the variables so that all inputs are of comparable magnitude. The temperature, humidity and pressure values are normalised by subtracting the mean and dividing by the standard deviation which is done separately at each height.

Figure 3 shows an example of training data for one instant in time with a total of 18,000 samples. The number 18,000 comes from the product of 80 LAMs, each with 25 coarse-graining regions each covering 9 different sizes (from 144km to 9km). In total the training data consists of 18,000 samples, 12 times a day, for 25 days, leading to 5.4 million samples for training and validation.

2.3 Neural network architecture

As an initial foray into using coarse-grained data for training machine-learned parameterizations, the architectures have been kept intentionally simple. A fully-connect neural network is used, as illustrated in Fig. 4. The input layer is a vector 215 elements long. This consists of temperature at each of the 70 vertical levels in the atmospheric
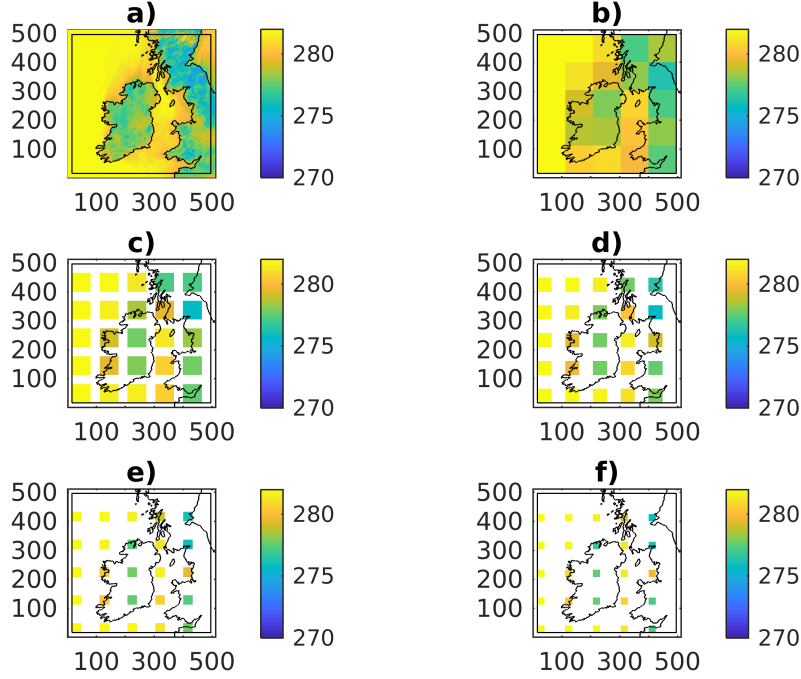
Figure 2. Example of coarse-graining the data in one limited-area model (LAM). a) original 1.5 km data, b) to f) coarse-graining by averaging over square regions which are 96, 64, 48, 32 and 24 gridboxes wide, respectively. This example is using near-surface temperature at 00Z on 1 January 2020 for the region centred on 54.26°N, 7°W located over Ireland and parts of the United Kingdom.

model and 70 levels of humidity and 70 levels of pressure catenated together. An additional 5 scalar values are then appended. These are the time-varying normalised mean-sea-level pressure and 4 constant-in-time ancillary values, namely the mean and standard deviation of the model orography, the fraction of the gridbox that is land and the normalised size of the coarse-graining region.

Specifically, the first 140 inputs to the neural network are height-dependent profiles of liquid-ice temperature T_{li} and total humidity q_T which are defined as:

$$T_{li} = T - \frac{L_v}{c_p}q_{cl} - \frac{L_f}{c_p}q_{cf} \tag{1}$$

and

$$q_T = q_v + q_{cl} + q_{cf} \tag{2}$$

where L_v and L_f are the latent heats of vaporization and fusion respectively, c_v is the specific heat capacity of air at constant pressure. The variables q_v, q_{cl} and q_{cf} are the height-dependent gridbox-mean specific humidity, gridbox-mean liquid water content (LWC) and gridbox-mean ice water content (IWC).

The output vector, the target, is a vector 210 elements long consisting of 70 levels each of in-cloud LWC (q_{cl}/C), in-cloud IWC (q_{cf}/C) and cloud fraction (C) all appended to each other.
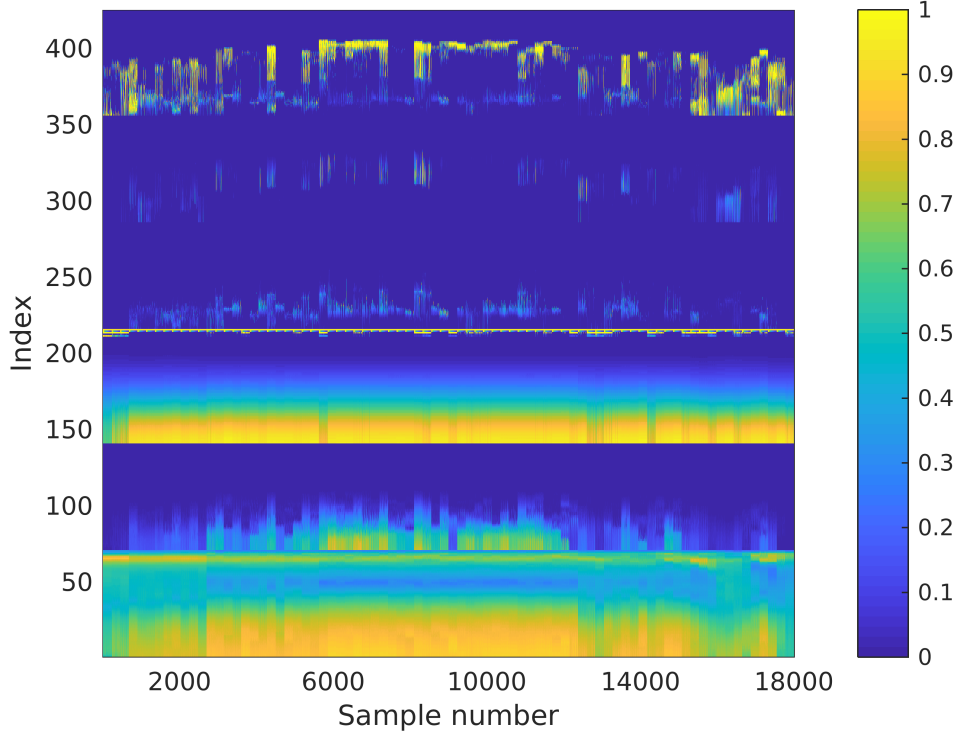
Figure 3. Example of the normalised training data for an instant in time. The x dimension has 18,000 samples, associated with spatial averaging over a 5×5 matrix, at 9 different resolutions, and over 80 domains. The y dimension is 425, corresponding to 3 sets of 70-element vectors for temperature, humidity and pressure at each height in the atmospheric model, followed by 5 scalar variables and then 3 sets of 70-element vectors for profiles of liquid water content, ice water content and bulk cloud fraction.

A parameter sweep is carried out to determine the optimum configuration of the hidden layers. In between the input and output layer, there are between 1 and 5 hidden layers, all consisting of either 256 or all 512 nodes each. In the hidden layers the activation function is a leaky rectified linear unit (Leaky ReLU, $y(x) = max(0.1x, x)$). The output layer is a capped ReLU ($y(x) = min(max(x, 0), 1)$). Note than when predicting liquid, q_{cl}, or ice water content, q_{cf}, the network is trained, and makes predictions, in terms of in-cloud condensate amount, i.e. q_{cl}/C and q_{cf}/C.

All the candidate networks were trained over 20 epochs, with an initial learning rate of $1.0e^{-4}$, reducing by a factor of 2 every 5 epochs. The network was trained using batches of 64 and the Adam optimiser. A two-dimensional grid-search algorithm was used to find the optimal set of layers and nodes. Each permutation was trained with 5 different initialisations and the median final mean-squared error (MSE) taken as indicative of the performance of that architecture. The evolution of the MSE for those medians for the training and validation data as a function of epoch is show in Fig. 5. The network using 4 hidden layers of 256 nodes was used for subsequent evaluation and coupling to the MetUM. This was chosen on balance by considering the low value of MSE and the number of trainable parameters.
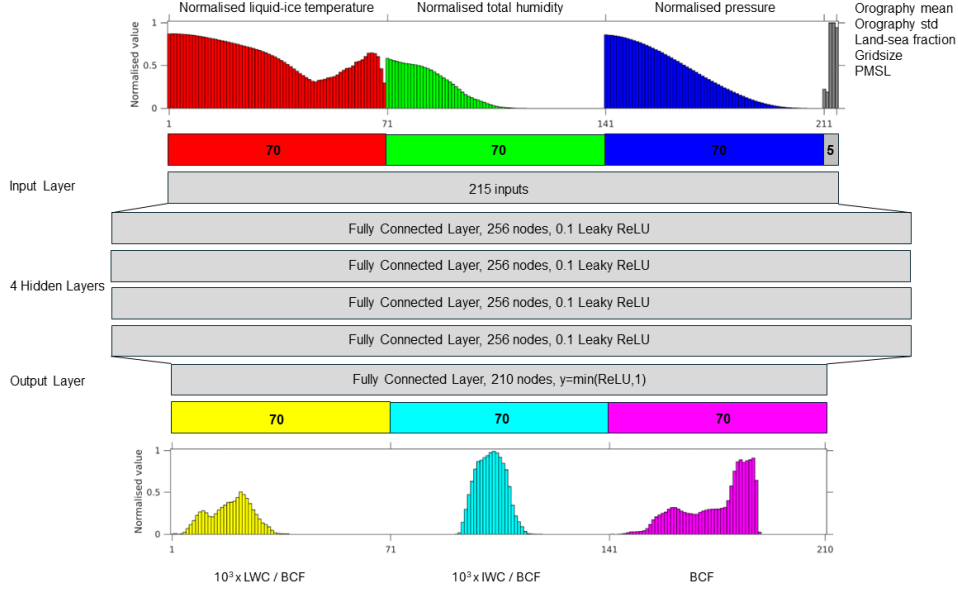
Figure 4. Schematic of the fully-connected neural network which consists of 215 inputs, 4 hidden layers of 256 nodes each and 210 outputs. An example of an input vector and associated output vector are given. The input vector consists of 215 normalised "features". This contains 70 values of temperature (red), 70 values of humidity (green), 70 levels of pressure (blue) and 5 scalars (gray) representing mean orography, standard deviation of orography, land fraction, grid-box size and pressure at mean-sea-level (PMSL). The output vector consists of 210 normalised "targets". This consists of 70 values of in-cloud liquid water content (yellow), 70 values of in-cloud ice water content (cyan) and 70 values of bulk cloud fraction (magenta).

An example of some of the ancillary input data is shown in Fig. 6. The mean orography, in metres, gets rescaled by dividing by 4000 m. The standard deviation of the sub-grid orography, in metres, gets rescaled by dividing by 1200 m. The land-fraction is already a dimensionless value with values between 0 and 1. The size of the coarse-graining region in kilometers gets rescaled by dividing by 144 km.

The types of surfaces that the dataset covers are is illustrated by Fig. 7, which shows a scatter plot of the normalised standard deviation of the sub-grid orography against the normalised gridbox mean orography. Sea points have both low altitude and are flat (low standard deviation of orography). By constrast, some locations are at higher altitude and flat, or high altitude and more rugged or low altitude and rugged.

Investigation into the distribution of the data as a function of coarse-grained land fraction found that around 69% of the training data is over the sea, 27% is over the land and 4% is coastal to some degree.

Initial development of some neural networks used the entirety of the training data and the training loss reduced with increasing epochs. However, when looking at the validation data set, it became clear that the network tended to always predict small values of cloud fractions. As a result, histograms of column-mean cloud fraction were produced for the training data. Figure 8 shows such a histogram for a single 18,000 sample file. Note that the model has 70 levels in the vertical, with around 50 of them below 20km. As a result, a completely cloudy tropical troposphere would give a column-
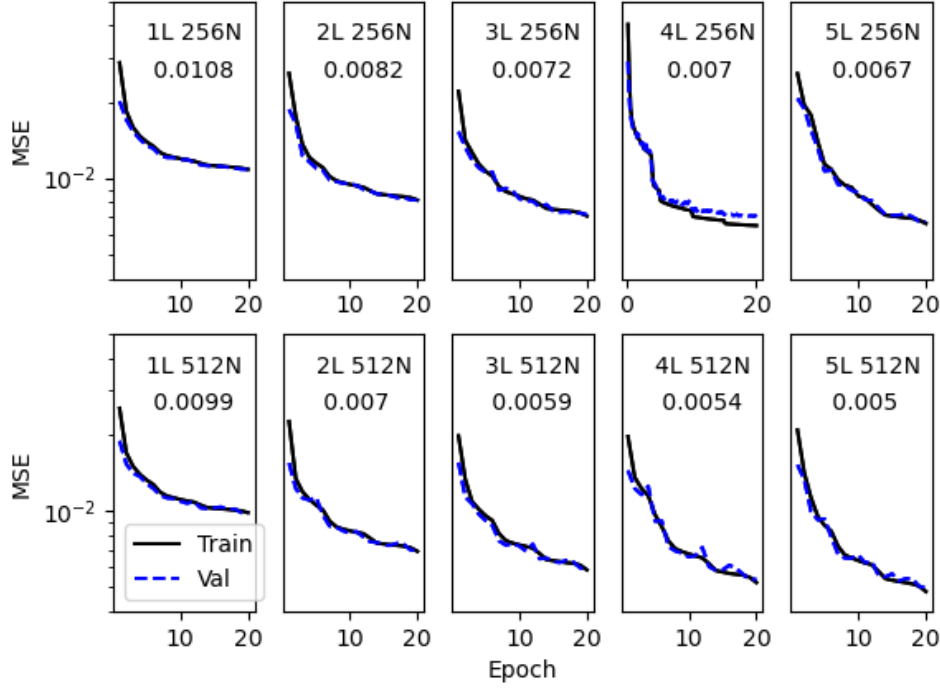
Figure 5. Training and validation curves showing the evolution of mean-squared error (MSE) as a function of epochs for multi-layer perceptrons with 1 to 5 hidden layers (L) each with 256 or 512 nodes (N). The MSE after 20 epochs appears on each panel.

mean cloud fraction of around 0.7, which explains why there are no occurrences of column-mean cloud fraction greater than this.

From Fig. 8, there is notable imbalance, with 65% of the data having a column-mean cloud fraction of 0.1 or less and only 0.2% of the data having a column-mean cloud fraction larger than 0.5. The initial models trained using these data were in effect hedging their bets. They predicted low values of cloud fraction, because that was an effective way of minimizing the cost function during training.

To overcome this issue, the training data was rebalanced. The histogram bin with 790 samples corresponding to column-mean cloud fraction of between 0.3 and 0.4 was chosen as an anchor point and all 790 samples were retained. For each of the bins with more samples, 790 randomly-selected samples were retained. For each of the bins with less samples, the data in that bin was shuffled and then replicated until the correct number of samples could be taken from that bin. This process of sub-sampling and over-sampling ensures a more balanced data set. The choice of which bin to use as an anchor is a compromise between massively reducing the total amount of training data, and having most of the training data being replicates. As shown in Fig. 8, using the 0.3 to 0.4 bin as the anchor point leads to $6 \times 790 = 4740$ samples being retained for this file. A reduction in the training data by this factor was typical across all the days and times, leading to a final training and validation dataset consisting of 1.4 million samples. With an 80:20 training:validation split, this produced 1.1 million samples for training.
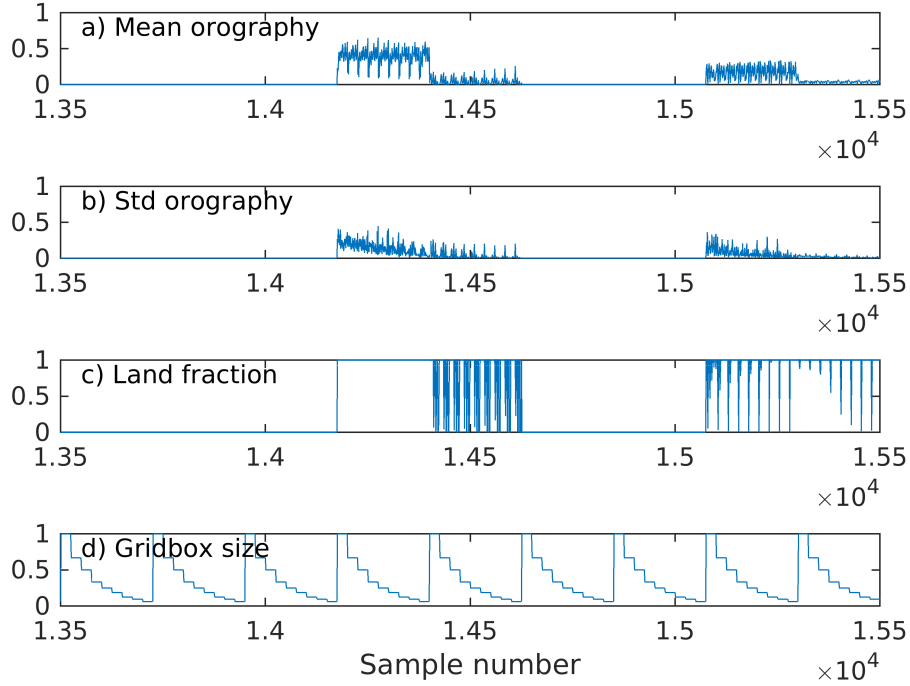
Figure 6. A sample of the ancillary data. A cross-section through Fig. 3 between samples 13500 and 15500, showing a) normalised mean orography, b) normalised standard deviation of orography, c) land fraction and d) normalised gridbox size.

The cost function calculates the average of the MSE from the profiles of bulk cloud fraction, gridbox mean liquid water content, gridbox mean ice water content and the maximum cloud fraction in each column.

2.4 Coupling of machine-learnt algorithm to the Unified Model

The machine-learning model was developed using Python, Keras (Chollet et al., 2015) and Tensorflow (Abadi et al., 2015) packages. This was coupled to the Unified Model (written in Fortran) as follows. Fortran routines were written to represent a multi-layer perceptron with a given architecture. Python code was then written to edit the Fortran routines and automatically paste the weights and biases from the trained model into data structures to make them available as Fortran arrays (Morcrette, 2024). This allowed the neural network trained in Python to be available for use within the Unified Model and to be compiled as another Fortran subroutine, with no dependence on external libraries or complex compilation.

In the Unified Model, the machine-learnt cloud scheme was implemented by first calculating profiles of T_{li} and q_T in each column. Information about the gridbox size is calculated as the square root of the gridbox area, itself inferred from the latitude and the known details of the latitude-longitude grid, i.e.

$$dx = \sqrt{\frac{2\pi R_E cos(\phi)}{nx} \times \frac{\pi R_E}{ny}} \tag{3}$$
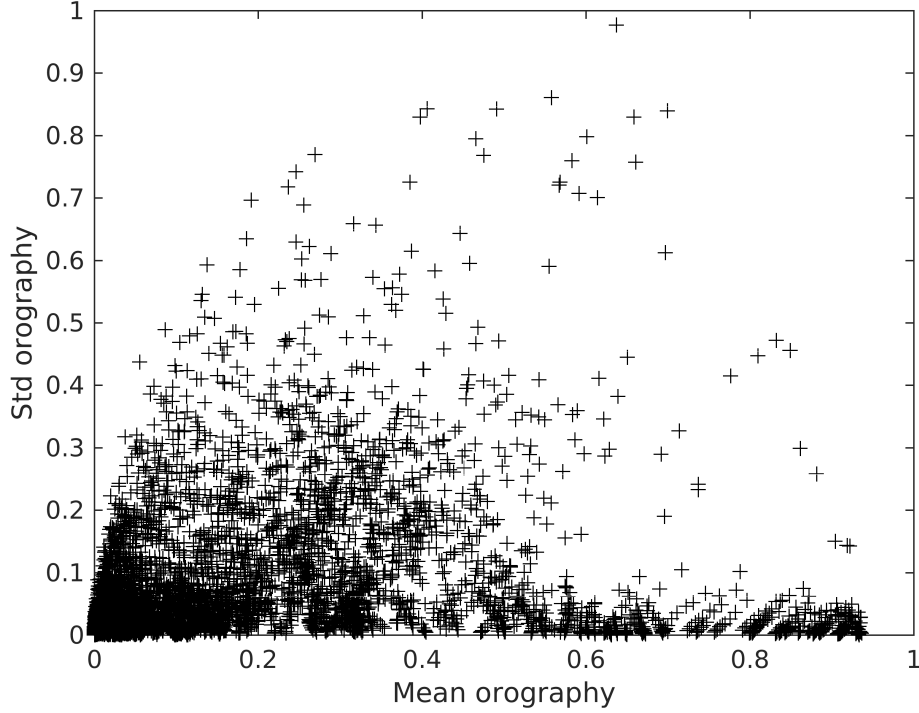
Figure 7. Scatter plot of normalised standard deviation of gridbox orography against normalised mean orography. The normalisation factors are 1200m for the standard deviation and 4000m for the mean.

where R_E is the radius of the Earth, ϕ is the latitude and nx and ny are the number of gridpoints in the east-west and north-south directions. Processing the high-resolution data to a range of sizes before training and using a normalised $dx/144$ as an input to the neural network means that scale-aware information is available, allowing the machine-learnt cloud scheme to be used in the Unified Model with its regular latitude-longitude but non-uniform spacing grid.

The normalised temperature, humidity, pressure and ancillary information is then passed to the neural network which predicts the profiles of in-cloud condensates and cloud fraction. The gridbox mean condensates are found by multiplying the in-cloud condensates by the cloud fraction and these are then used as inputs to the radiation and microphysics schemes. The new water vapor specific humidity is found by subtracting the diagnosed gridbox mean condensates from the total humidity calculated just prior to calling the neural network and the (dry) temperature T is found from T_{li} and the newly diagnosed condensate amounts. The updated temperature and humidity fields are then evolved as normal by the model dynamics and other physics packages.

3 Results

3.1 Evaluation of trained model

The off-line performance of the machine-learnt algorithm was evaluated by comparing its predictions to true values from the end of the month-long simulations, and 5 days after the end of the training data, i.e. 12 UTC 31 January 2020.
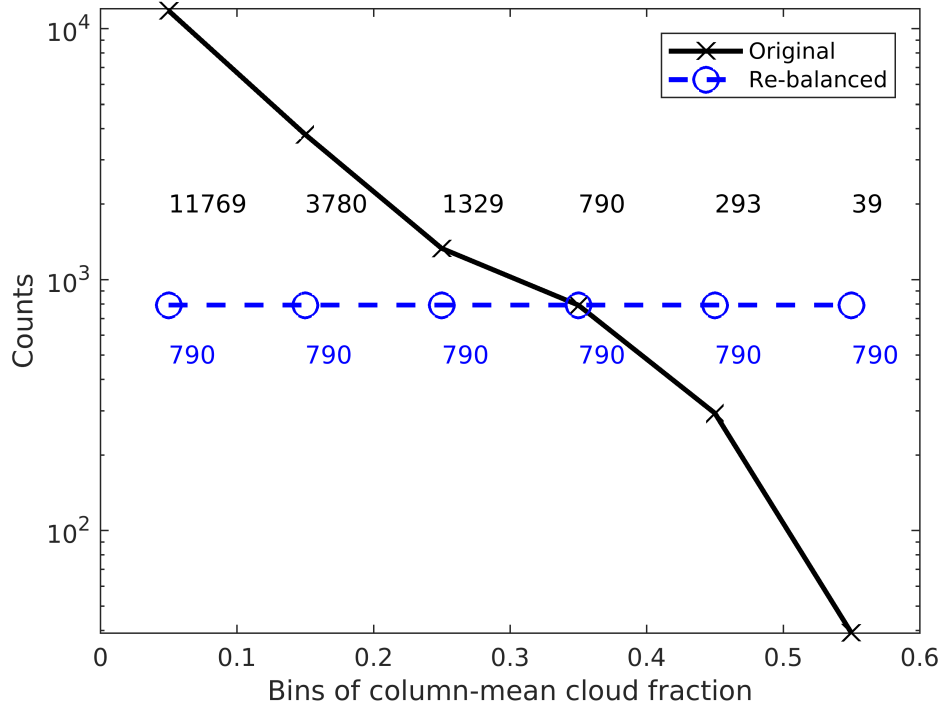
Figure 8. Histogram for a single snapshot in time, showing the distribution of column-mean cloud fraction over 18,000 samples (black). The rebalanced data set, using sub-sampling and over-sampling to enforce an identical number of samples in each bin is also shown (blue). Note the logarithmic y-scale
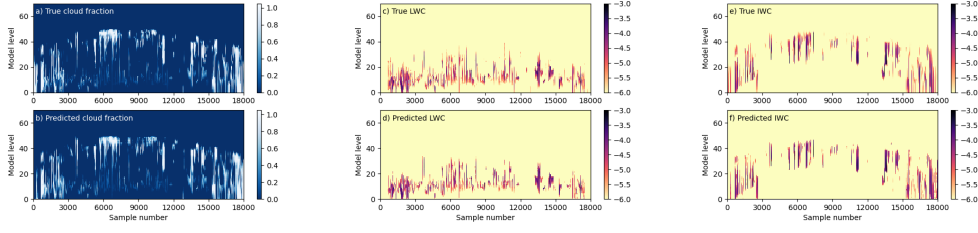
Figure 9 shows a vertical cross-section through the test data. There are 18,000 samples, consisting of 225 samples from each of the 80 LAMs, and for each LAM there are 25 samples at each of 9 progressively finer scales. The LAMs are ordered by latitude, from south to north, and hence the plots are pseudo latitude-height cross sections. The machine-learnt model predicts profiles of cloud fraction, with regional variations in cloud base, cloud top and cloud fraction being reproduced. In terms of gridbox-mean condensate amounts and their phase the neural network captures the condensate amounts that vary over several orders of magnitude.

Figure 10 shows two-dimensional histograms of the predicted and true values of cloud fraction and condensate amounts. The cloud fraction histogram does not have all the points perfectly along the diagonal, but does capture the clear and overcast cases fairly well. In terms of the condensate amounts, the distribution is fairly tight around the diagonal, but is more dispersed at lower condensate amounts.

Figure 11 shows that on average the neural network does a good job of capturing the profiles of cloud fraction and liquid and ice condensate amounts.

### 3.2 Climate model simulation

Several climate simulation experiments were carried out. These are detailed in Table 1 and based on the GA8 configuration (Willett et al., 2024). Each simulation was run for 5 years using a latitude-longitude gridlength of $1.25° \times 1.875°$.

Figure 9. Vertical cross-sections through the test data showing a,c,e) true and b,d,f) predicted value of a,b) cloud fraction, c,d) log_{10} of gridbox-mean liquid water content [kg/kg] and e,f) log_{10} of gridbox-mean ice water content [kg/kg].
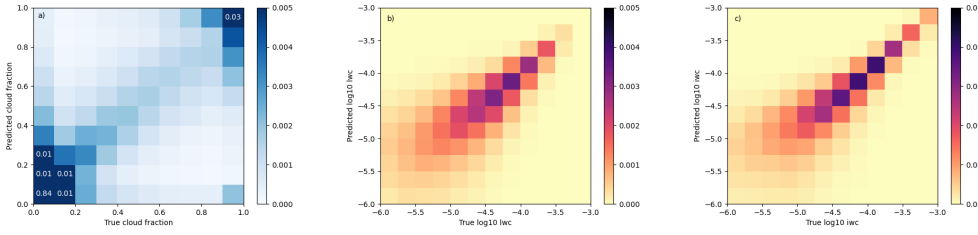
Figure 10. Normalised two-dimensional histograms comparing the predicted and true values of a) cloud fraction b) liquid water content and c) ice water content. The color scale has been clipped, but any values larger than that are labelled.
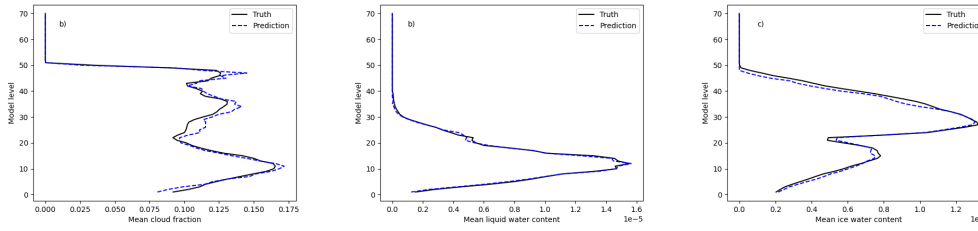
Figure 11. Profiles of mean true and predicted cloud fraction from the test data.

The new parametrizaton, developed from cloud-resolving model machine learning (CRMML), is referred to as "Caramel". Each experimemt includes the Caramel cloud scheme coupled interactively to the atmospheric model with it producing cloud cover and condensate amounts used as inputs to both the radiation and microphysics schemes. These schemes in turn affect the temperature and humidity structures in each atmospheric column and affect the time evolution of the model.

The control climate model includes the PC2 cloud cover and cloud condensate scheme (Wilson, Bushell, Kerr-Munslow, Price, & Morcrette, 2008; Wilson, Bushell, Kerr-Munslow, Price, Morcrette, & Bodas-Salcedo, 2008). The initial implementations of this scheme (Walters et al., 2011, 2014, 2017) assumed that all clouds were represented by the scheme. During subsequent development, it was found that PC2 missed some of the cloudiness associated with convection. As a result, the GA8 control simulation used here includes a small amount of extra diagnostic cloud to represent "convective cores" (Walters et al., 2019).

The Caramel scheme was trained on data containing all types of clouds and as such it is expected to account for all cloudiness. Several simulations have been carried out. The first experiment, Caramel 1, simply replaces the PC2 cloud scheme with the machine-learnt one. The second, Caramel 2, uses the machine-learnt cloud scheme but, to avoid double counting some convective cloudiness, does not include the radiative impact of the convective cores which was added to GA8. This configuration of GA8, with the machine-learnt cloud scheme replacing PC2, and the cores switched off, was then built upon and further sensitivity tests were carried out.

The machine-learnt scheme predicts a cloud cover in each gridbox, known as a bulk cloud fraction, which represents clouds that can be either liquid, ice or mixed phase. The new scheme also predicts liquid and ice water contents. The microphysics and radiation schemes however need to know how much of the cloud cover is purely liquid, purely ice and how much is mixed phase. A first assumption is to consider the case of minimum overlap between the phases; the liquid and ice fraction are given by the respective condensate-weighted portion of the bulk fraction. This is referred to an overlap of zero. The alternate assumption is that liquid and ice condensate co-exist as much as possible. This is referred to an overlap of one. In the code, both liquid and ice cloud fraction are calculated using both assumptions and a equal blend of the two approaches is used, which is referred to as an overlap of a half. The impact of changing the overlap assumption from 0.5 to 0.0 or 1.0 is summarised in Table 1.

In the development of their offline machine-learnt cloud scheme, Henn et al. (2024) found it beneficial to filter out any predicted cloud fraction less than 6.5% in order to improve their radiative fluxes. Inspired by their result, a simulation where any cloud fraction predicted as less than 6.5% leads to the cloud fraction and condensate amounts being set to zero was carried out.

The control simulation (GA8, Willett et al. 2024) represents the radiative impact of sub-grid variability in condensate by using the implementation of the Monte Carlo independent column approximation (McICA) described by (Hill et al., 2011) and a parameterization of cloud condensate fractional standard deviation (FSD) described by (Hill et al., 2015)). Initial tests of the machine-learnt cloud scheme with the cores switched off showed too much reflected short-wave. A final simulation, with the FSD increased by a third, was carried out as a result. This last simulation has global top-of-atmosphere short-wave and long-wave radiation values which are quite close to the observations.

The development of a global atmospheric model leads to the generation of a multitude of different plots and metrics. As a simple measure of the impact of the machine-learnt cloud scheme, Table 1 gives global area-weighted mean of outgoing short-wave and long-wave radiation for different configurations.

Our preferred configuration is given in the last row of Table 1. This consists of the machine-learnt cloud scheme replacing PC2, with the convective cores switched off, a mixed-phase overlap of 0.5 and a rescaling of the FSD by a factor of 1.33. This configuration, referred to simply as Caramel, is the simulation for which further diagnostics and maps will be presented.

Figure 12 shows that the climate simulation with the interactive machine-learnt cloud parameterization has only a limited impact on the top-of-atmosphere (TOA) outgoing long-wave radiation (OLR) with root-mean square (RMS) very similar between the experiment and the control. In terms of TOA outgoing short-wave radiation (OSR), the picture is less favourable (Fig. 13). Most of the tropical oceans are brighter, apart from the regions of stratocumulus off the coasts of California, Peru and Namibia. The distribution of global rainfall has a slightly larger RMSE with the neural-network cloud scheme (Fig. 14)

Table 1. Description of the various 5-year climate model simulations and global area-weighted outgoing short-wave (OSR) and outgoing long-wave radiation (OLR).

Short name	Description					OSR (Wm^{-2})	OLR (Wm^{-2})
Obs	EBAF4.1					99.00	240.30
Control	GA8					101.32	239.19
Caramel 1	GA8	+MLCloud		Overlap=0.5		109.53	235.73
Caramel 2	GA8	+MLCloud	−cores	Overlap=0.5		107.70	236.14
Caramel 3	GA8	"	"	Overlap=0.0		108.33	236.36
Caramel 4	GA8	"	"	Overlap=1.0		107.70	236.57
Caramel 5	GA8	"	"	Overlap=0.5	Henn	107.54	236.03
Caramel	GA8	+MLCloud	−cores	Overlap=0.5	FSD×1.33	98.14	239.25

Table 2. Global-mean radiation diagnostics calculated relative to CERES EBAF (NASA/LARC/SD/ASDC, 2019). All units are Wm^{-2}.

Variable	Observations	GA8 Control	MLCloud Experiment
Outgoing SW TOA	99.0	101.32	98.14
Clear sky outgoing SW TOA	53.7	56.12	55.72
Outgoing LW TOA	240.3	239.19	239.25
Clear sky outgoing LW TOA	266.1	262.01	260.10
Absorbed SW	241.3	239.07	242.25
Planetary albedo at TOA	29.1	29.77	28.83
Clear-sky albedo at TOA	15.8	16.49	16.37
SW cloud radiative forcing	-45.3	-45.20	-42.42
LW cloud radiative forcing	25.8	22.82	20.85
Net downward rad. at TOA	1.0	-0.12	3.00
Net CRE at TOA	-19.5	-22.37	-21.58
Net surface shortwave	163.8	166.00	168.92
Clear-sky downward surface shortwave	241.0	244.17	244.64
Clear-sky upward surface shortwave	29.1	30.65	30.71
Total downward surface shortwave	187.0	190.68	194.24
Net downward LW = -1. * upward LW	-40 to -63	-57.74	-60.00
Surface downward LW	345.4	338.72	337.02
Clear-sky surface downward LW	317.7	313.00	314.11
Upward SW	23.2	24.68	25.32
Net surface radiation (downward)	110.3	108.26	108.92

Table 3. Additional global-mean diagnostics.

| Variable	Units	Observations	GA8 Control	MLCloud Experiment
Total cloud amount	-		0.6	0.67	0.76
Total precipitation	$mm\ day^{-1}$	2.61-2.695	3.00	2.94
Evaporation	$mm\ day^{-1}$	2.61-2.695	3.00	2.94
Sensible heat flux	Wm^{-2}	17-25	20.55	20.01
Latent heat flux	Wm^{-2}	78-86	86.96	85.07
MSLP	hPa	1011.1	1011.99	1011.96
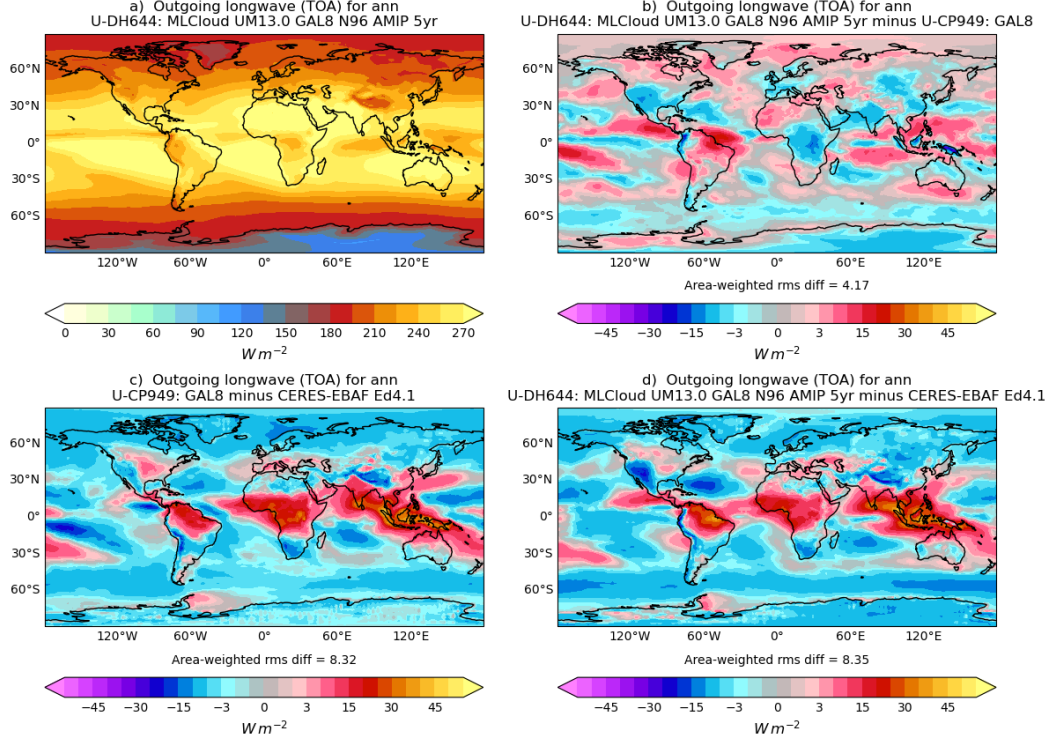
Figure 12. Map of a) outgoing long-wave radiation in the new MLCloud simulation (configuration in last row of Table 1). Maps of differences between b) the new MLCloud simulation and the GA8 control, c) the control and the observations and d) the new MLCloud simulation and the observations.

4 Discussion

The Met Office Unified Model (MetUM) is a seamless modeling system which is used for both climate and weather modelling (Brown et al., 2012) and that will remain the case with the next generation of atmospheric model (Adams et al., 2019).

While the whole-forecast emulators described by Bi et al. (2023) or Lam et al. (2023) propose a radically different approach to global weather forecasting, this work has focused on the development of machine-learnt cloud parameterization for inclusion into a climate model. Now that the scheme is available within the MetUM code it could also be tested and developed in weather forecasting mode. Recent work by Husain et al. (2024) describes an interesting hybrid approach given the better performance of a weather emulator at the large-scale but the better performance of the traditional model at the smaller scale. They describe running a weather emulator before a traditional model and the physics-based model is then nudged towards the weather emulator. As a result, it is still reasonable to develop improvements to the traditional model's parameterization schemes, and this can be via the use of coarse-grained high-resolution simulations as described here.

A novel aspect of this work is that the neural network is trained to predict profiles of cloud cover, in-cloud liquid water content and in-cloud ice water content all in one go. The prediction of in-cloud condensate amounts was found necessary as a way of dealing with potential inconsistencies between cloud cover and condensate which should never be zero if the other is non-zero.
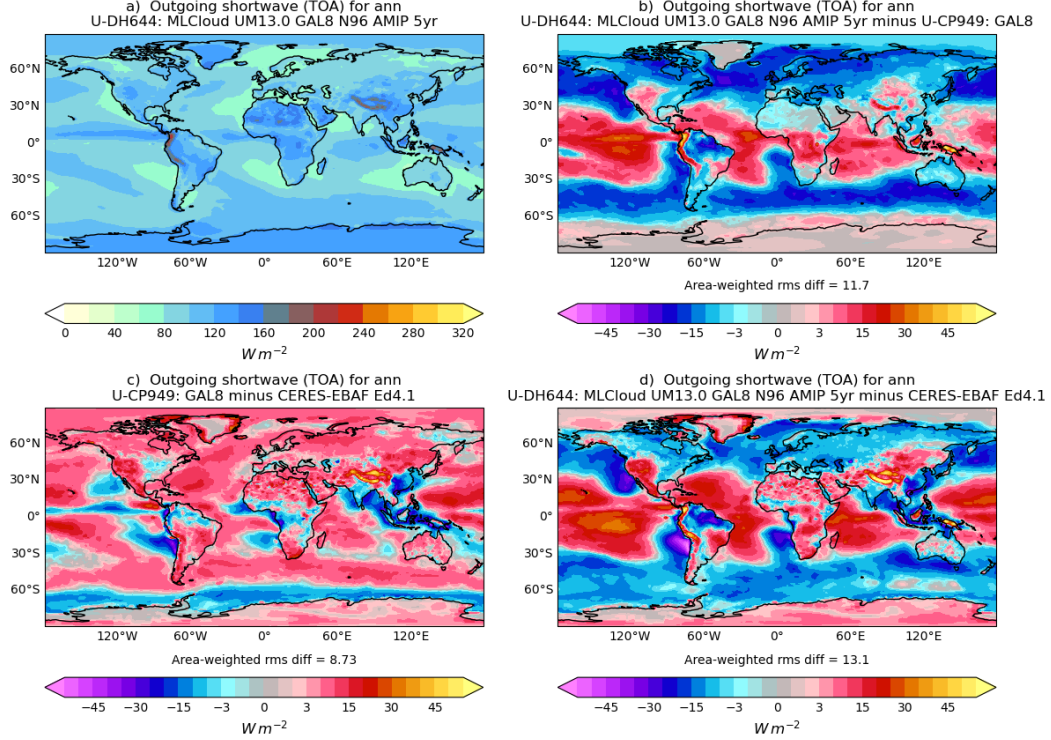
Figure 13. Same as 12 but for outgoing short-wave radiation.

When implemented in the climate model, the various global mean metrics and global maps of cloud and radiation fields show that the machine-learning scheme can be used to run stable multi-year climate simulations. The metrics however show a slight deterioration compared to the control. This is not altogether surprising, as the implementation of new parameterization schemes often necessitates the retuning of other parameterization schemes (Hourdin et al., 2013; Walters et al., 2019).

The current machine-learnt cloud scheme could also be improved. Avenues for future work include:

1. extending the training dataset to include more data from different seasons and years,
2. using data from a global cloud-resolving model or increasing the number of high-resolution domains, possibly focusing on regions where the impact of the neural network on the predicted cloud-radiative effects could be improved,
3. better ways of dealing with the imbalanced dataset (there are two orders of magnitude more examples of columns with small cloud amounts than columns with large cloud amounts), this could be through rebalancing the dataset before training or by accounting for it as part of the cost function,
4. inclusion of more physics into the cost-function and using some of the physical constraints recommended by Grundner et al. (2024),
5. training using climate-invariant inputs such as relative humidity instead of specific humidity as recommended by Beucler et al. (2024),
6. using a one-dimensional convolutional neural network to better, and more efficiently, capture the importance of vertical gradients in temperature and humidity when predicting cloud properties,
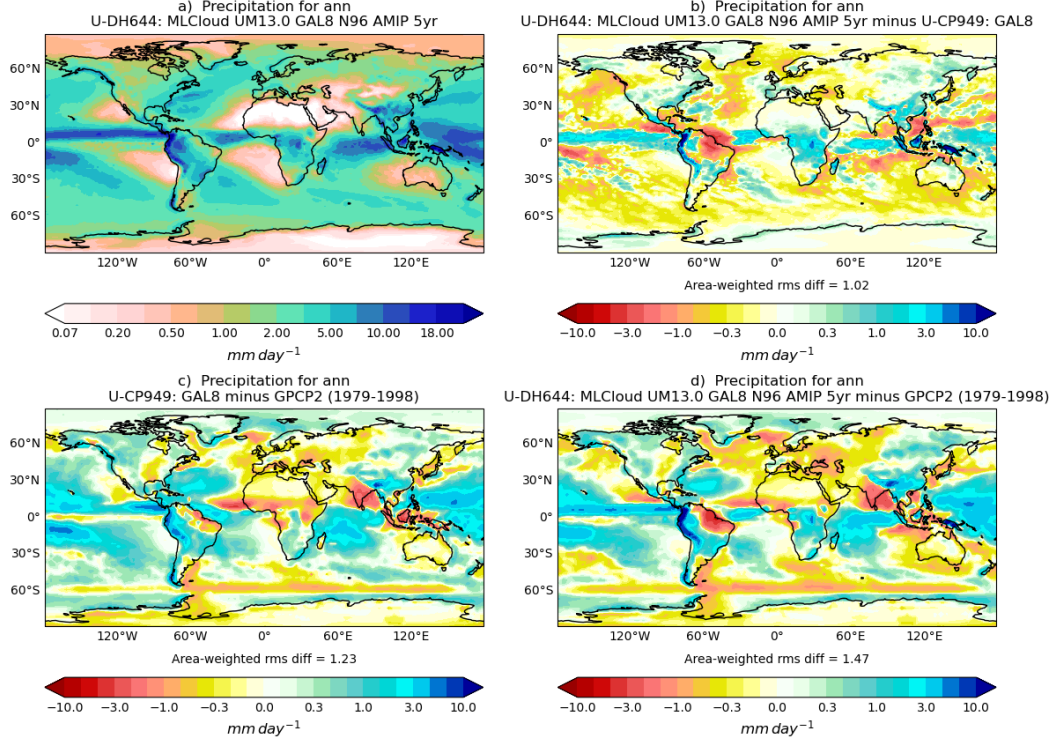
Figure 14. Same as 12 but for precipitation.

7. including profiles of the 2 components of the horizontal wind as inputs to the algorithm, in order to capture the importance of shear in increasing cloudiness laterally within a gridbox,

8. including profiles of the coarse-grained vertical component of the wind as an input to represent the importance of large-scale ascent,

9. developing a hurdle model approach, a first model predicts one of two classes: completely clear column or some cloud present somewhere. Then a second model, trained only on the cloudy columns, predicts what the profiles of cloud cover and condensates would be.

5 Conclusions

A series of high-resolution simulations, carried out using an atmospheric model normally used for operational regional weather forecasting, have been coarse-grained and used as training data for a neural network that predicts profiles of cloud cover as well as profiles of liquid and ice water contents. The inputs are the profiles of temperature, humidity and pressure as well as information about the land-sea mask, size of the coarse-graining region and the mean and standard deviation of the orography. No information from the control cloud scheme is used as an input to the neural network. This machine-learnt cloud scheme is coupled to a global climate model. This includes the radiation and microphysics parameterization schemes using the machine-learnt cloud fraction and liquid and ice water contents as inputs to radiative transfer and precipitation production calculations. The coupling of the machine-learnt cloud scheme to the rest of the model also occurs through latent heating. When clouds form or dissipate through condensation or evaporation, the change in phase has an impact on the vertical temperature struc-

ture. This fully coupled and interactive machine-learnt cloud scheme is used to run several multi-year interactive simulations.

Although the cloud-radiative properties of the new climate simulations show that some further development is required, this paper provides a new example of how machine learning can be used to introduce information gleaned from coarse-grained high-resolution simulations into a global atmospheric model. The manner in which the machine-learning model is coupled to the general circulation model will be of broad interest to the community due to the abundance of future applications, covering potential improvements to any parameterization within the atmospheric, land, cryosphere and ocean environmental system.

Acknowledgments

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/` (Software available from tensorflow.org)

Adams, S., Ford, R., Hambley, M., Hobson, J., Kavčič, I., Maynard, C., . . . Wong, R. (2019). LFRic: Meeting the challenges of scalability and performance portability in weather and climate models. *Journal of Parallel and Distributed Computing*, *132*, 383-396. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0743731518305306` doi: https://doi.org/10.1016/j.jpdc.2019.02.007

Beucler, T., Gentine, P., Yuval, J., Gupta, A., Peng, L., Lin, J., . . . Pritchard, M. (2024). Climate-invariant machine learning. *Science Advances*, *10*(6), eadj7250. Retrieved from `https://www.science.org/doi/abs/10.1126/sciadv.adj7250` doi: 10.1126/sciadv.adj7250

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3d neural networks. *Nture*, *619*, 533-538. doi: 10.1038/s41586-023-06185-3

Bony, S., Stevens, B., Frierson, D. M. W., Jakob, C., Kageyama, M., Pincus, R., . . . Webb, M. J. (2015). Clouds, circulation and climate sensitivity. *Nature Geoscience*, *8*(4), 261–268. (doi: 10.1038/ngeo2398)

Brenowitz, N. D., & Bretherton, C. S. (2018). Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, *45*(12), 6289-6298. (doi: 10.1029/2018GL078510)

Brenowitz, N. D., & Bretherton, C. S. (2019). Spatially extended tests of a neural network parametrization trained by coarse-graining. *J. Adv. Model. Earth Syst.*, *11*(8), 2728-2744. (doi: 10.1029/2019MS001711)

Brown, A., Milton, S., Cullen, M., Golding, B., Mitchell, J., & Shelly, A. (2012).

Unified modeling and prediction of weather and climate: a 25 year journey. *Bull. Amer. Meteor. Soc.*, *93*, 1865-1877. (doi:10.1175/BAMS-D-12-00018.1)

Bush, M., Boutle, I., Edwards, J., Finnenkoetter, A., Franklin, C., Hanley, K., . . . Weeks, M. (2023). The second met office unified model–jules regional atmosphere and land configuration, ral2. *Geoscientific Model Development*, *16*(6), 1713–1734. Retrieved from `https://gmd.copernicus.org/articles/16/1713/2023/` doi: 10.5194/gmd-16-1713-2023

Chantry, M., Hatfield, S., Dueben, P., Polichtchouk, I., & Palmer, T. (2021). Machine learning emulation of gravity wave drag in numerical weather forecasting. *Journal of Advances in Modeling Earth Systems*, *13*(7), e2021MS002477. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021MS002477` (e2021MS002477 2021MS002477) doi: https://doi.org/10.1029/2021MS002477

Chevallier, F., Chéruy, F., Scott, N. A., & Chédin, A. (1998). A neural network approach for a fast and accurate computation of longwave radiative budget. *J. Appl. Meteor*, *37*, 1385–1397.

Chollet, F., et al. (2015). *Keras.* `https://keras.io`.

de Burgh-Day, C. O., & Leeuwenburg, T. (2023). Machine learning for numerical weather and climate modelling: a review. *Geoscientific Model Development*, *16*(22), 6433–6477. Retrieved from `https://gmd.copernicus.org/articles/16/6433/2023/` doi: 10.5194/gmd-16-6433-2023

Donlon, C. J., Martin, M., Stark, J. D., Roberts-Jones, J., Fiedler, E., & Wimmer, W. (2012). The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. *Remote Sensing of Environment*, *116*, 140-158. (doi:10.1016/j.rse.2010.10.017)

Dueben, P. D., & Bauer, P. (2018). Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development*, *11*(10), 3999–4009. Retrieved from `https://gmd.copernicus.org/articles/11/3999/2018/` doi: 10.5194/gmd-11-3999-2018

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., & Yacalis, G. (2018). Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, *45*(11), 5742-5751. (doi: 10.1029/2018GL078202)

Golaz, J.-C., Larson, V. E., & Cotton, W. R. (2002). A PDF-based model for boundary layer clouds. Part I: Method and model description. *J. Atmos. Sci.*, *59*, 3540-3551.

Grundner, A., Beucler, T., Gentine, P., & Eyring, V. (2024). Data-driven equation discovery of a cloud cover parameterization. *Journal of Advances in Modeling Earth Systems*, *16*(3), e2023MS003763. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023MS003763` (e2023MS003763 2023MS003763) doi: https://doi.org/10.1029/2023MS003763

Grundner, A., Beucler, T., Gentine, P., Iglesias-Suarez, F., Giorgetta, M. A., & Eyring, V. (2022). Deep learning based cloud cover parameterization for ICON. *J. Adv. Model. Earth Syst.*, *14*(12). (https://doi.org/10.1029/2021MS002959)

Henn, B., Jauregui, Y. R., Clark, S. K., Brenowitz, N. D., McGibbon, J., Watt-Meyer, O., . . . Bretherton, C. S. (2024). A machine learning parameterization of clouds in a coarse-resolution climate model for unbiased radiation. *Journal of Advances in Modeling Earth Systems*, *16*(3), e2023MS003949. doi: https://doi.org/10.1029/2023MS003949

Hill, P. G., Manners, J., & Petch, J. C. (2011). Reducing noise associated with the Monte Carlo Independent Column Approximation for weather forecasting models. *Quart. J. Roy. Meteor. Soc.*, *137*, 219-228. (doi:10.1002/qj.732)

Hill, P. G., Morcrette, C. J., & Boutle, I. A. (2015). A regime-dependent parametrization of subgrid-scale cloud water content variability. *Quart. J.*

Roy. Meteor. Soc., *141*, 1975-1986. (doi:10.1002/qj.2506)

Hourdin, F., Grandpeix, J.-Y., Rio, C., Bony, S., Jam, A., Cheruy, F., . . . Roehrig, R. (2013). LMDZ5B: the atmospheric component of the IPSL climate model with revisited parameterizations for clouds and convection. *Clim. Dyn*, *40*, 2193–2222. (doi:10.1007/s00382-012-1343-y)

Husain, S. Z., Separovic, L., Caron, J.-F., Aider, R., Buehner, M., Chamberland, S., . . . Zadra, A. (2024). *Leveraging data-driven weather models for improving numerical weather prediction skill through large-scale spectral nudging.* Retrieved from https://arxiv.org/pdf/2407.06100

Illingworth, A. J., Hogan, R. J., O'Connor, E. J., Bouniol, D., Brooks, M. E., Delanoe, J., . . . Wrench, C. L. (2007). CloudNet - continuous evaluation of cloud profiles in seven operational models using ground-based observations. *Bull. Amer. Meteor. Soc.*, *88*, 883-898.

Jakob, C., & Klein, S. A. (1999). The role of vertically varying cloud fraction in the parametrization of microphysical processes in the ECMWF model. *Quart. J. Roy. Meteor. Soc.*, *125*, 941-965.

Kain, J. S., Willington, S., Clark, A. J., Weiss, S. J., Weeks, M., Jirak, I. L., . . . Suri, D. (2017). Collaborative efforts between the United States and United Kingdom to advance prediction of high-impact weather. *Bull. Amer. Meteor. Soc.*, *98*, 937-948. (doi: 10.1175/BAMS-D-15-00199.1)

Keat, W. J., Stein, T. H. M., Phaduli, E., Landman, S., Becker, E., Bopape, M. M., . . . Webster, S. (2019). Convective initiation and storm life cycles in convection-permitting simulations of the Met Office Unified Model over South Africa. *Quart. J. Roy. Meteor. Soc.*, *145*(721), 1323-1336. (doi: 10.1002/qj.3487)

Keisler, R. (2022). *Forecasting global weather with graph neural networks.* Retrieved from https://doi.org/10.48550/arXiv.2202.07575

Khairoutdinov, M., Randall, D., & DeMott, C. (2005). Simulations of the atmospheric general circulation using a cloud-resolving model as a superparameterization of physical processes. *J. Atmos. Sci.*, *62*, 2136-2154. (doi: 10.1175/JAS3453.1)

Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., . . . Hoyer, S. (2024). Neural general circulation models for weather and climate. *Nature*, *tbc*(tbc), tbc. doi: 10.1038/s41586-024-07744-y

Kooperman, G. J., Pritchard, M. S., Burt, M. A., Branson, M. D., & Randall, D. A. (2016). Robust effects of cloud superparameterization on simulated daily rainfall intensity statistics across multiple versions of the community earth system model. *Journal of Advances in Modeling Earth Systems*, *8*(1), 140-165. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015MS000574 doi: https://doi.org/10.1002/2015MS000574

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., & Belochitski, A. A. (2013). Using ensemble of neural networks to learn stochastic convection parameterizations for climate and numerical weather prediction models from data simulated by a cloud resolving model. *Advances in Artificial Neural Systems*, *2013*(1), 485913. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1155/2013/485913 doi: https://doi.org/10.1155/2013/485913

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. *Monthly Weather Review*, *133*(5), 1370 - 1383. Retrieved from https://journals.ametsoc.org/view/journals/mwre/133/5/mwr2923.1.xml doi: 10.1175/MWR2923.1

Kumar, S., Mukhopadhyay, P., & Balaji, C. (2024). A machine learning based deep convective trigger for climate models. *Clim. Dyn.*. doi: https://doi.org/10.1007/s00382-024-07332-w

Lam, R., gaSanchez Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., . . . Battaglia, P. (2023). Learning skillful medium-range global weather forecasting. *Science*, *382*(6677), 1416-1421. doi: 10.1126/science.adi2336

Lean, H. W., Clark, P. A., Dixon, M., Roberts, N. M., Fitch, A., Forbes, R., & Halliwell, C. (2008). Characteristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom. *Mon. Wea. Rev.*, *136*, 3408-3424. (doi:10.1175/2008MWR2332.1)

Le Treut, H., & Li, Z.-X. (1991). Sensitivity of an atmospheric general circulation model to presribed SST changes: feedback effects associated with the simulation of cloud optical properties. *Clim. Dyn.*, *5*, 175-187.

Lewis, H., Bowyer, J., Broad, A., Chamberlain-Clay, A., Jones, C., Chan, S., . . . Morcrette, C. (2022). Using machine learning to find cloud base height: a didactic challenge. *Weather*, *77*(11), 391-395. (DOI: 10.1002/wea.4163)

Morcrette, C. J. (2024). *ENNUF: Easy Neural Networks for Use in Fortran*. Retrieved from `https://zenodo.org/records/12581295`

Morcrette, C. J., O'Connor, E. J., & Petch, J. C. (2012). Evaluation of two cloud parametrization schemes using ARM and Cloud-Net observations. *Quart. J. Roy. Meteor. Soc.*, *138*, 964-979. (doi:10.1002/qj.969)

NASA/LARC/SD/ASDC. (2019, 6 12). *CERES energy balanced and filled (ebaf) TOA and surface monthly means data in NetCDF edition 4.1*. NASA Langley Atmospheric Science Data Center DAAC. Retrieved from `https://doi.org/10.5067/TERRA-AQUA/CERES/EBAF_L3B.004.1`

Orr, A., Phillips, T., Webster, S., Elvidge, A., Weeks, M., Hosking, S., & Turner, J. (2013). Met Office Unified Model high-resolution simulations of a strong wind event in Antarctica. *Quart. J. Roy. Meteor. Soc.*, *140*(684), 2287-2297. (doi: 10.1002/qj.2296)

Randall, D. (1989). Cloud parameterization for climate modeling: Status and prospects. *Atmos. Res.*, *23*, 345-361.

Randall, D., Khairoutdinov, M., Arakawa, A., & Grabowski, W. (2003). Breaking the cloud parameterization deadlock. *Bull. Amer. Meteor. Soc.*, *84*, 1547-1564.

Rasp, S., Pritchard, M., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, *115*, 9684-9689. (doi: 10.1073/pnas.1810286115)

Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophysical Research Letters*, *44*(24), 12396-12417. (doi: 10.1002/2017GL076101)

Seifert, A., & Rasp, S. (2020). Potential and limitations of machine learning for modeling warm-rain cloud microphysical processes. *Journal of Advances in Modeling Earth Systems*, *12*(12), e2020MS002301. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002301` (e2020MS002301 10.1029/2020MS002301) doi: https://doi.org/10.1029/2020MS002301

Slingo, J. (1987). The development and verification of a cloud prediction scheme for the ECMWF model. *Quart. J. Roy. Meteor. Soc.*, *113*, 899-927.

Smith, R. N. B. (1990). A scheme for predicting layer cloud and their water content in a general circulation model. *Quart. J. Roy. Meteor. Soc.*, *116*, 435-460.

Sommeria, G., & Deardorff, J. W. (1977). Subgrid-scale condensation in models of non-precipitating clouds. *J. Atmos. Sci.*, *34*, 344-355.

Song, H.-J., Lim, B., & Joo, S. (2019). Evaluation of rainfall forecasts with heavy rain types in the high-resolution Unified Model over South korea. *Wea. Forecasting*, *34*, 1277-1293. (doi: 10.1175/WAF-D-18-0140.1)

Stein, T. H. M., Parker, D. J., Hogan, R. J., Birch, C. E., Holloway, C. E., Lister, G. M. S., . . . Woolnough, S. J. (2015). The representation of the West African

monsoon vertical cloud structure in the Met Office Unified Model: an evaluation with CloudSat. *Quart. J. Roy. Meteor. Soc.*, *141*(693), 3312-3324. (doi: 10.1002/qj.2614)

Stensrud, D. J. (2007). *Parameterization schemes: Keys to understanding numerical weather prediction models.* Cambridge University Press.

Sundqvist, H. (1978). A parametrization scheme for non-convective condensation including prediction of cloud water content. *Quart. J. Roy. Meteor. Soc.*, *104*, 677-690.

Tiedtke, M. (1993). Representation of clouds in large-scale models. *Mon. Wea. Rev.*, *121*, 3040-3061.

Tompkins, A. (2002). A prognostic parameterization for the subgrid-scale variability of water vapor and clouds in large-scale models and its use to diagnose cloud cover. *J. Atmos. Sci.*, *59*, 1917-1942.

Ukkonen, P., & Mäkelä, A. (2019). Evaluation of machine learning classifiers for predicting deep convection. *Journal of Advances in Modeling Earth Systems*, *11*(6), 1784-1802. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001561 doi: https://doi.org/10.1029/2018MS001561

Van Weverberg, K., Boutle, I. A., Morcrette, C. J., & Newsom, R. K. (2016). Towards retrieving critical relative humidity from ground-based remote-sensing observations. *Quart. J. Roy. Meteor. Soc.*, *142*, 2867-2881. (doi: 10.1002/qj.2874)

Van Weverberg, K., & Morcrette, C. (2022). Sensitivity of cloud-radiative effects to cloud fraction parametrizations in tropical, mid-latitude and arctic kilometre-scale simulations. *Quart. J. Roy. Meteor. Soc.*, *148*(746), 2563-2586. (DOI: 10.1002/qj.4325)

Van Weverberg, K., Morcrette, C., & Boutle, I. (2021). A bimodal diagnostic cloud fraction parameterization. Part II: Evaluation and resolution sensitivity. *Mon. Wea. Rev.*, 859–878. (https://doi.org/10.1175/MWR-D-20-0230.1)

Walters, D. N., Baran, A., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., . . . Zerroukat, M. (2019). The Met Office Unified Model Global Atmosphere 7.0 and JULES Global Land 7.0 configurations. *Geosci. Model Dev.*, *12*, 1909–1963. (doi: 10.5194/gmd-2017-291)

Walters, D. N., Best, M. J., Bushell, A. C., Copsey, D., Edwards, J. M., Falloon, P. D., . . . Williams, K. D. (2011). The Met Office Unified Model Global Atmosphere 3.0/3.1 and JULES Global Land 3.0/3.1 configurations. *Geosci. Model Dev.*, *4*, 919-941. (doi:10.5194/gmd-4-919-2011)

Walters, D. N., Brooks, M., Boutle, I., Melvin, T., Stratton, R., Vosper, S., . . . Xavier, P. (2017). The Met Office Unified Model Global Atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations. *Geosci. Model Dev.*, *10*, 1487-1520. (doi.org/10.5194/gmd-10-1487-2017)

Walters, D. N., Williams, K. D., Boutle, I. A., Bushell, A. C., Edwards, J. M., Field, P. R., . . . Vidale, P. L. (2014). The Met Office Unified Model Global Atmosphere 4.0 and JULES Global Land 4.0 configurations. *Geosci. Model Dev.*, *7*, 361-386. (doi:10.5194/gmd-7-361-2014)

Watanabe, M., Emori, S., Satoh, M., & Miura, H. (2009). A pdf-based hybrid prognostic cloud scheme for general circulation model. *Clim. Dyn.*, *33*, 795-816. (doi:10.1007/s00382-008-0489-0)

Webster, S., Uddstrom, M., Oliver, H., & Vosper, S. (2008). A high resolution modelling case study of a severe weather event over New Zealand. *Atmos. Sci. Let.*, *9*(3), 119-128. (doi: 10.1002/asl.172)

Weyn, J. A., Durran, D. R., & Caruana, R. (2020). Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *J. Adv. Model. Earth Syst.*, *12*(9). (https://doi.org/10.1029/2020MS002109)

Willett et al.   (2024).   *The Met Office Unified Model Global Atmosphere 8.0 configuration.*   Retrieved from `https://cms.ncas.ac.uk/unified-model/configurations/ga8.0-gc4.0/`

Wilson, D. R., Bushell, A. C., Kerr-Munslow, A. M., Price, J. D., & Morcrette, C. J. (2008).   PC2: A prognostic cloud fraction and condensation scheme. I: Scheme description. *Quart. J. Roy. Meteor. Soc.*, *134*, 2093-2107.

Wilson, D. R., Bushell, A. C., Kerr-Munslow, A. M., Price, J. D., Morcrette, C. J., & Bodas-Salcedo, A. (2008). PC2: A prognostic cloud fraction and condensation scheme. II: Climate model simulations.   *Quart. J. Roy. Meteor. Soc.*, *134*, 2109-2125.

Xu, K.-M., & Randall, D. A.   (1996).   A semiempirical cloudiness parameterization for use in climate models. *J. Atmos. Sci.*, *53*(21), 3084–3102.