

# Improving vertical detail in simulated temperature and humidity data using machine learning

Joana D. da Silva Rodrigues<sup>1</sup>  | Cyril J. Morcrette<sup>1,2</sup> 

<sup>1</sup>Met Office, Exeter, UK

<sup>2</sup>Department of Mathematics and Statistics, University of Exeter, Exeter, UK

## Correspondence

Joana D. da Silva Rodrigues, Met Office, FitzRoy Road, Exeter, EX1 3PB, UK.  
Email: [joana.rodrigues@metoffice.gov.uk](mailto:joana.rodrigues@metoffice.gov.uk)

## Funding information

Met Office; AI4PEX; UK Research and Innovation (UKRI); Horizon Europe, Grant/Award Numbers: 10114295, 10103109, 10093450

## Abstract

Atmospheric models used for weather forecasting and climate predictions discretise the atmosphere onto a vertical grid. There are however atmospheric phenomena that occur on scales smaller than the thickness of those model layers. The formation of low-level clouds due to temperature inversions is an example. This leads to atmospheric models underestimating, or even missing, these clouds and their radiative effects. Using radiosonde observations as training data, a machine learning model is used to improve the vertical detail of modelled profiles of temperature and specific humidity. In addition, a physics-informed machine learning model is developed and compared to the traditional approach; showing improvements in the cloud fraction profiles calculated from its predictions. The vertically enhanced profiles also improve the representation of layers of convective inhibition and anomalous refractivity gradients. This work facilitates targeted improvements to the representation of certain atmospheric processes without the burden of increased memory and computational cost from increasing vertical resolution throughout the whole model.

## KEYWORDS

clouds, convolutional neural network, machine learning, parameterisation, radiosondes, super resolution

## 1 | INTRODUCTION

Atmospheric models are useful tools for predicting the weather and making projections of the Earth's future climate (Bauer et al., 2015; Brown et al., 2012; Williams et al., 2020). It is not possible, however, to run these models with high enough resolution to accurately capture all the processes that are important to the phenomena we wish to predict. Atmospheric models discretise the atmosphere onto a regular horizontal grid and parameterisation schemes are employed to represent the net effect of subgrid processes on the resolved scale

(Stensrud, 2007). The atmosphere is also discretised in the vertical, with information about the meteorological variables stored on a finite set of vertical levels. There are several atmospheric phenomena which can occur over vertical depths which are shallow relative to the model grid. This includes thin clouds and layers of convective inhibition (CIN) and regions of anomalous refractivity gradients.

Thin clouds that form in the marine boundary layer are known to not be well represented in many climate models. This leads to a large uncertainty in their radiative feedbacks (Bony & Dufresne, 2005; Slingo, 1990). From a

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 Crown copyright. *Atmospheric Science Letters* published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society. This article is published with the permission of the Controller of HMSO and the King's Printer for Scotland.

weather forecasting point of view, accurate forecasts of the presence of clouds are important for aviation safety and for reliable predictions of the amount of solar radiation reaching the surface (e.g. Gultepe et al., 2019; Mendes et al., 2024).

To improve cloud fraction predictions, some schemes have been developed to sharpen the profiles of temperature ( $T$ ) and humidity ( $q$ ) near inversions and hence get a better profile of relative humidity (RH) on a temporary high-resolution (HR) grid (Boutle & Morcrette, 2010). When this higher-resolution thermodynamic data is passed to the cloud cover scheme (e.g. Smith, 1990), it can improve predictions of cloud cover. Another approach is to use information about the vertical gradients of  $T$  and  $q$  to infer  $\text{RH}_{\text{crit}}$ , the critical RH at which cloud formation first occurs (Van Weverberg et al., 2016) and to use that when calculating the cloud fraction.

Meanwhile, thin layers of CIN can hinder the development of convection in the atmosphere. Some weather forecasting and climate models include a consideration of CIN in their convective parametrisation scheme (Fletcher & Bretherton, 2010; Kuang & Bretherton, 2006), but this can be an issue if the CIN itself is poorly represented due to the relatively coarse vertical grid of the model (Stirling & Stratton, 2012). In addition, strong vertical gradients of temperature and humidity will affect atmospheric refractivity and radio-wave propagation which can result in radar ducting, a phenomenon which affects meteorological and aeronautical radar surveillance systems (Wang et al., 2012).

While reducing the vertical thickness of the model layers is likely to improve the representation of these atmospheric processes, the increase in resolution comes with a significant increase in cost. This extra computational cost comes from both the extra memory required to store all the atmospheric variables on the higher resolution grid and from the extra calculations that need to be done on all this data throughout the entirety of the model.

A pragmatic solution is proposed. The bulk of the model structure remains as it is, but finer vertical detail is used where it can provide the most benefit. For example, normal resolution thermodynamic profiles are refined before calculating cloud fraction. The HR cloud fraction is then being regridded to the normal grid before proceeding with the rest of the model integration. This study focuses on the first step in this process: the development of techniques to sharpen the details in vertical profiles of atmospheric  $T$  and  $q$ .

Super resolution (SR) is a branch of image processing that consists in enhancing the quality (resolution) of an image. It is also a field that has been gaining from advancements in machine learning (ML) in the past

years, with the ML SR models showing success when compared to traditional methods. These models have already been used for downscaling problems in weather and climate science. Chen et al. (2019), for example, trained generative adversarial networks (GANs) to improve the resolution of weather radar echo images and Stengel et al. (2020) developed GANs and convolutional neural networks (CNNs) capable of super-resolving solar irradiance and wind velocity climate data. In addition, Wu et al. (2021) compared the performance of CNNs and GANs for the purpose of emulating urban climate information, and a GAN originally built by Wang et al. (2021) for image SR was successfully used by Manepalli et al. (2020) to downscale wind maps.

Most examples of ML models for the purpose of SR in meteorological applications are either two- or three dimensional. The  $T$  and  $q$  data we consider are one-dimensional, so ML techniques adapted from the field of audio super-resolution are used to learn how to enhance the vertical resolution in an atmospheric model thermodynamic profile. Observations from weather balloons are used as a source of HR training data. These are coarse-grained to the resolution of the atmospheric model and a neural network is trained to reintroduce the detail that is lost as a result. It is shown that this improves the fine details in the structure of the thermodynamic profiles as well as having the potential to improve the diagnosis of cloud fraction and CIN regions. The implementation of the scheme into a weather or climate model, along with the inevitable tuning and subsequent evaluation of its impact, will be the subject of future work.

The paper is structured as follows. The training data and ML techniques are described in Section 2. Results of the evaluation of the vertical-resolution enhancement technique in terms of fidelity in reconstructed  $T$ ,  $q$  and cloud fraction are presented in Section 3. A discussion and prospects for future implementation in an atmospheric model appear in Section 4. Conclusions are drawn in Section 5.

## 2 | METHOD

### 2.1 | Data

The data used in this project were obtained from the Atmospheric Radiation Measurement (ARM) archive, a U.S. Department of Energy Office of Science user facility managed by the Biological and Environmental Research Program. The data were originally obtained from ARM radiosonde launches and interpolated to a fixed time-height grid, for each atmospheric state variable (Jensen et al., 2013). The main reason for using these

radiosonde data is their high vertical resolution in comparison to our atmospheric models.

Data from four different locations around the globe were used for training: Alaska, Australia, Azores and Oklahoma. The data were divided such that one full year was used for training (01/10/2013–30/09/2014), 6 weeks for validation (01/10/2014–15/11/2014) and another 6 weeks for testing (16/11/2014–31/12/2014), which gives an ideal split of 80%/10%/10%.

Furthermore, to evaluate how the model performed at a location and time different from those it was trained on, data from Ascension Island, located in the South Atlantic Ocean, were chosen for a period from 01/06/2016–15/07/2016.

Using a complete year of data for training, as well as data from locations with very different climates, means a wide range of values of  $T$  and  $q$  is covered, which should help avoid extrapolation and improve generalisability.

Using data from different locations however leads to a complication, as they are not on the same grid in the vertical. As a first step, all observational data are interpolated to a HR 50-m grid up to a height of 6400 m (128 levels).

For the input to the machine-learning model, the same radiosonde data were coarse-grained to Unified Model (UM) resolution. This Low Resolution (LR) grid has 29 levels up to 6 km, considerably less than the HR grid.

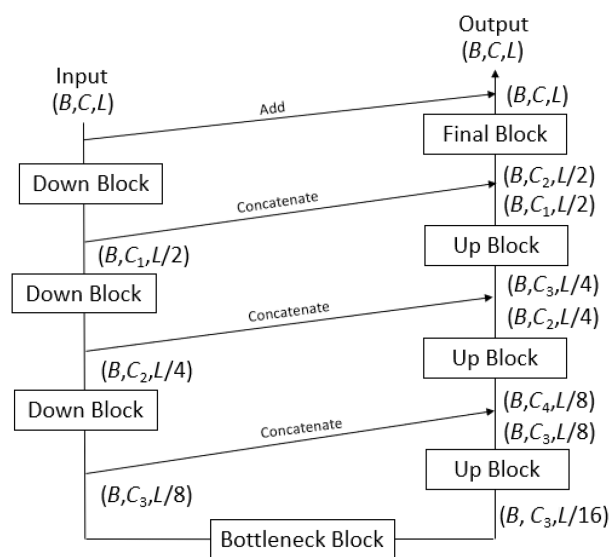
Model data are not used as the input. If it were, then biases in the model predictions could affect the results and the machine-learning model would be trying to both compensate for model errors and enhance the vertical resolution. By using observations throughout, it is ensured that the only difference between the input and output is their resolution.

This way, when the loss is being reduced during training or certain metrics are measured during testing, it is absolutely certain that only the error coming from the profiles' resolution is being taken into account and nothing more.

## 2.2 | Model

A CNN with a U-Net architecture was built based on audio SR work done by Kuleshov et al. (2017). The model was adapted to take two channels as input, corresponding to the LR  $T$  and  $q$  profiles, and output back those two same profiles at HR. A graphical representation of its structure is shown in Figure 1.

The model is a *pre-upsampling* SR model, meaning that the first step is to bring the input to HR space before performing convolutions on it. This helps reduce the



**FIGURE 1** Graphical representation of the structure of the one-dimensional super resolution convolutional neural network.  $B$  is the size of the batch, that is the number of samples,  $C$  is number of channels and  $L$  is the length of the input.

upsampling artefacts. As such, before being passed to the ML model, a cubic interpolation is performed on the LR input.

On the ‘down’ side of the U-net, convolutions are performed on the data, followed by a leaky Rectified Linear Unit with a slope of 0.2 (Kuleshov et al., 2017). At every block the number of channels increases while the length of the data is reduced (halved). At the ‘bottleneck’ of the model, the data is  $2^4$  times smaller than the input.

The opposite happens on the ‘up’ side, where every block contains a convolutional layer, a Rectified Linear Unit and a pixel shuffle layer (Shi et al., 2016), which increases the data in length and reduces it in number of channels. The pixel shuffle layer has been shown to produce less artefacts than the common upsample layer (Odena et al., 2016).

Each of these blocks is followed by a skip connection where the output from that block is stacked to the output from the corresponding block on the down side, which has the same dimensions. These connections have been proved to help stabilise training as they ensure that the spatial information of the features is not lost (Orhan & Pitkow, 2017). Dropout and batch normalisation are also introduced in this part of the model as they help avoid overfitting and make the model more stable.

At the final layer of the neural network there is no activation function, dropout or batch normalisation. Furthermore, instead of stacking the cubic interpolated input and output together, they are added to each other. This operation at the end means that the network does not

need to learn all the patterns in the profiles of  $T$  and  $q$ , it just needs to learn a correction. This makes sense for cases like this one, in which the output is very similar to the input, and it helps speed up training.

Using a grid search, the optimal hyperparameters for the model were found and the final version of the model had a total of about 300,000 trainable parameters. The model was trained using an Adam optimiser, with a learning rate of  $10^{-4}$ , following a minibatches approach (each containing 64 samples). It was trained for a total of 100 epochs but using early stopping, to help avoid overfitting. The loss function used for training was a simple Mean Squared Error (MSE) between the model's predictions and the target.

### 2.3 | Physics-informed ML

Physics-informed ML (PML) is the integration of physical knowledge into a ML model in order to ensure its predictions conform to known laws of physics. It has also been shown to lead to better generalisability, to less data being needed, faster training and an increase in interpretability and transparency in the models.

This can be done by adding a penalty to the cost function while training (*soft constraint*) or by changing the architecture of the model itself (*hard constraint*). A second version of the model was then trained to analyse and evaluate the effect of adding physics information to it, that version has exactly the same architecture as stated before but was trained with a penalised loss function.

Since the goal was not only to improve the profiles of  $T$  and  $q$  but also to do so in a physically consistent way, RH was used to penalise the CNN. During training, RH was calculated from the target  $T$  and  $q$  (and linearly interpolated pressure) as well as from the predictions of  $T$  and  $q$  for every batch, and the MSE of those two RH profiles was calculated and used as a penalty. A relative weight was attributed to the RH penalty to control how much it impacted the training of the CNN. Its optimal value was found to be 10%, with the remainder of the loss function being made up of equal parts normalised  $T$  and  $q$  MSE. Equation (1) describes the loss function used for the PML ( $L_{\text{PML}}$ ).

$$L_{\text{PML}} = (1 - \alpha) L_{\text{ML}} + \alpha P, \quad (1)$$

where  $L_{\text{ML}} = \text{MSE}(y_{\text{pred}}, y_{\text{truth}})$  is the traditional ML cost function (simple MSE),  $P = \text{MSE}(\text{RH}_{\text{pred}}, \text{RH}_{\text{truth}})$  is the RH penalty and  $\alpha$  is the weight attributed to that penalty.

To summarise, two versions of the CNN were trained, using the same architecture and training hyperparameters. The first one, a traditional ML model, was trained

with a simple MSE loss function based on  $T$  and  $q$ . The second one, a PML model, has a RH penalty contributing to the loss function.

## 3 | RESULTS

The test dataset, containing independent samples not used during training, was used to evaluate the models. The LR profiles are given to the previously trained model to get predictions of HR temperature and humidity, which are then compared against the input and the target profiles.

There are many ways of calculating cloud fraction from atmospheric soundings. As a way of demonstrating the impact of the refined  $T$  and  $q$  on potential cloud fraction calculations, the function presented by Wood and Field (2000) was used in this paper. This simple function is applicable to low-level cloud and can be used to quantify how vertical refinements to  $T$  and  $q$  affect the profile of RH and hence cloud fraction predictions. This function is described in Equation (2), where  $C$  is Cloud Fraction,  $q_{\text{T}}$  is total water content and  $q_{\text{s}}$  is the saturation specific humidity.  $A$  and  $B$  are constants set to 17.0 and 0.95, respectively.

$$C = \frac{1}{2} \{1 + \tanh(A(\overline{q_{\text{T}}}/\overline{q_{\text{s}}}) - B)\}. \quad (2)$$

Cloud-base height was then also calculated from the profiles of cloud fraction, defined as the lowest height at which cloud fraction is larger than 3/8. Finally, Convective Available Potential Energy (CAPE) and CIN were also obtained from the thermodynamical profiles using the relevant function from the MetPy library (May et al., 2022).

To measure and compare quantitatively the overall accuracy of the models' predictions, Root MSE (RMSE) is calculated for the predicted profiles in relation to the target. Tables 1–6 show the results for the two ML models, traditional and PML, and the cubic interpolation benchmark, for data from all five available locations separately (same number of samples for each), for temperature, specific humidity, cloud fraction, cloud-base height, CIN and CAPE, respectively. The results presented here are an average obtained by getting the results of the model for the testing set over five independent rounds of training for each version of the model.

Furthermore, Figures 2 and 3 show profiles predicted by the traditional ML model, compared against the cubic interpolated profiles and the HR ones that we are trying to emulate, for temperature and specific humidity, respectively.

**TABLE 1** RMSE of temperature (K) for the two models and the cubic, when compared against the target, at every location (for the testing dataset).

Location	Cubic	ML	PML
Azores	$4.254 \times 10^{-1}$	$(1.963 \pm 0.009) \times 10^{-1}$	$(2.037 \pm 0.081) \times 10^{-1}$
Australia	$3.520 \times 10^{-1}$	$(1.578 \pm 0.012) \times 10^{-1}$	$(1.649 \pm 0.054) \times 10^{-1}$
Alaska	$4.006 \times 10^{-1}$	$(1.758 \pm 0.038) \times 10^{-1}$	$(1.967 \pm 0.151) \times 10^{-1}$
Oklahoma	$4.408 \times 10^{-1}$	$(2.072 \pm 0.016) \times 10^{-1}$	$(2.190 \pm 0.097) \times 10^{-1}$
Ascension Island (2016)	$4.255 \times 10^{-1}$	$(2.435 \pm 0.023) \times 10^{-1}$	$(2.542 \pm 0.076) \times 10^{-1}$

**TABLE 2** RMSE of specific humidity ( $\text{kg kg}^{-1}$ ) for the two models and the cubic, when compared against the target, at every location (for the testing dataset).

Location	Cubic	ML	PML
Azores	$1.718 \times 10^{-4}$	$(1.630 \pm 0.005) \times 10^{-4}$	$(1.638 \pm 0.012) \times 10^{-4}$
Australia	$3.156 \times 10^{-4}$	$(2.980 \pm 0.006) \times 10^{-4}$	$(2.994 \pm 0.013) \times 10^{-4}$
Alaska	$4.557 \times 10^{-5}$	$(4.671 \pm 0.065) \times 10^{-5}$	$(4.517 \pm 0.018) \times 10^{-5}$
Oklahoma	$1.086 \times 10^{-4}$	$(1.068 \pm 0.002) \times 10^{-4}$	$(1.063 \pm 0.002) \times 10^{-4}$
Ascension Island (2016)	$2.876 \times 10^{-4}$	$(2.869 \pm 0.054) \times 10^{-4}$	$(2.863 \pm 0.032) \times 10^{-4}$

**TABLE 3** RMSE of cloud fraction for the two models and the cubic, when compared against the target, at every location (for the testing dataset).

Location	Cubic	ML	PML
Azores	$6.012 \times 10^{-2}$	$(5.488 \pm 0.094) \times 10^{-2}$	$(5.471 \pm 0.089) \times 10^{-2}$
Australia	$3.264 \times 10^{-2}$	$(3.271 \pm 0.085) \times 10^{-2}$	$(3.174 \pm 0.098) \times 10^{-2}$
Alaska	$1.021 \times 10^{-1}$	$(1.203 \pm 0.095) \times 10^{-1}$	$(0.991 \pm 0.026) \times 10^{-1}$
Oklahoma	$6.892 \times 10^{-2}$	$(6.612 \pm 0.096) \times 10^{-2}$	$(6.589 \pm 0.109) \times 10^{-2}$
Ascension Island (2016)	$3.998 \times 10^{-2}$	$(3.445 \pm 0.052) \times 10^{-2}$	$(3.442 \pm 0.117) \times 10^{-2}$

**TABLE 4** RMSE of cloud-base height (m) for the two models and the cubic, when compared against the target, at every location (for the testing dataset).

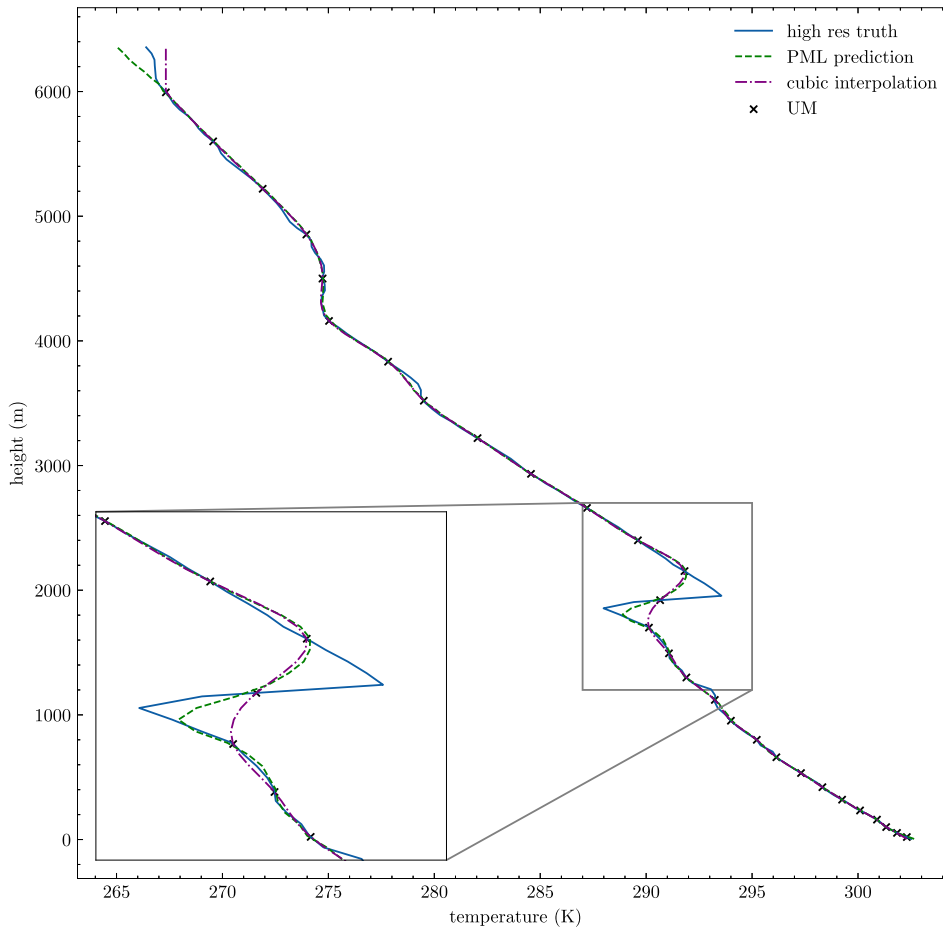
Location	Cubic	ML	PML
Azores	1240.515	$1217.106 \pm 39.364$	$1135.226 \pm 38.566$
Australia	1113.495	$1199.355 \pm 31.242$	$1163.647 \pm 33.452$
Alaska	510.368	$575.461 \pm 134.604$	$488.809 \pm 140.990$
Oklahoma	606.533	$725.955 \pm 26.498$	$688.757 \pm 25.128$
Ascension Island (2016)	1322.735	$1355.040 \pm 35.798$	$1291.993 \pm 33.832$

**TABLE 5** RMSE of CIN for the two models and the cubic, when compared against the target, at every location (for the testing dataset).

Location	Cubic	ML	PML
Azores	7.009	$3.177 \pm 0.354$	$3.336 \pm 0.180$
Australia	9.320	$4.683 \pm 0.660$	$5.556 \pm 0.843$
Alaska	0.370	$0.585 \pm 0.116$	$0.654 \pm 0.138$
Oklahoma	7.897	$3.870 \pm 1.777$	$4.824 \pm 2.153$
Ascension Island (2016)	24.664	$17.461 \pm 1.195$	$18.505 \pm 1.990$

**TABLE 6** RMSE of CAPE ( $\text{J}\cdot\text{kg}^{-1}$ ) for the two models and the cubic, when compared against the target, at every location (for the testing dataset).

Location	Cubic	ML	PML
Azores	49.128	$27.900 \pm 2.264$	$30.316 \pm 2.166$
Australia	187.333	$138.036 \pm 6.325$	$149.986 \pm 10.820$
Alaska	$1.831 \times 10^{-2}$	$(2.653 \pm 0.657) \times 10^{-2}$	$(2.419 \pm 0.463) \times 10^{-2}$
Oklahoma	2.436	$1.050 \pm 0.052$	$1.159 \pm 0.129$
Ascension Island (2016)	67.200	$36.804 \pm 3.694$	$37.088 \pm 4.019$



**FIGURE 2** Example of a temperature profile from the Tropical Western Pacific Central Facility, Darwin, Australia (18/11/2014, 22:30:30) predicted by the ML model (dashed green line) compared against the cubic interpolated profile (dash-dotted purple line) and the truth (full blue line). Values on the atmospheric model grid are shown by crosses. The left panel shows the profile up to 6 km and the right one zooms in on the inversion.

In Figure 4, an example of a cloud fraction profile resultant from PML predictions of  $T$  and  $q$  is shown, alongside the hr target and the profile calculated from cubic interpolated  $T$  and  $q$ .

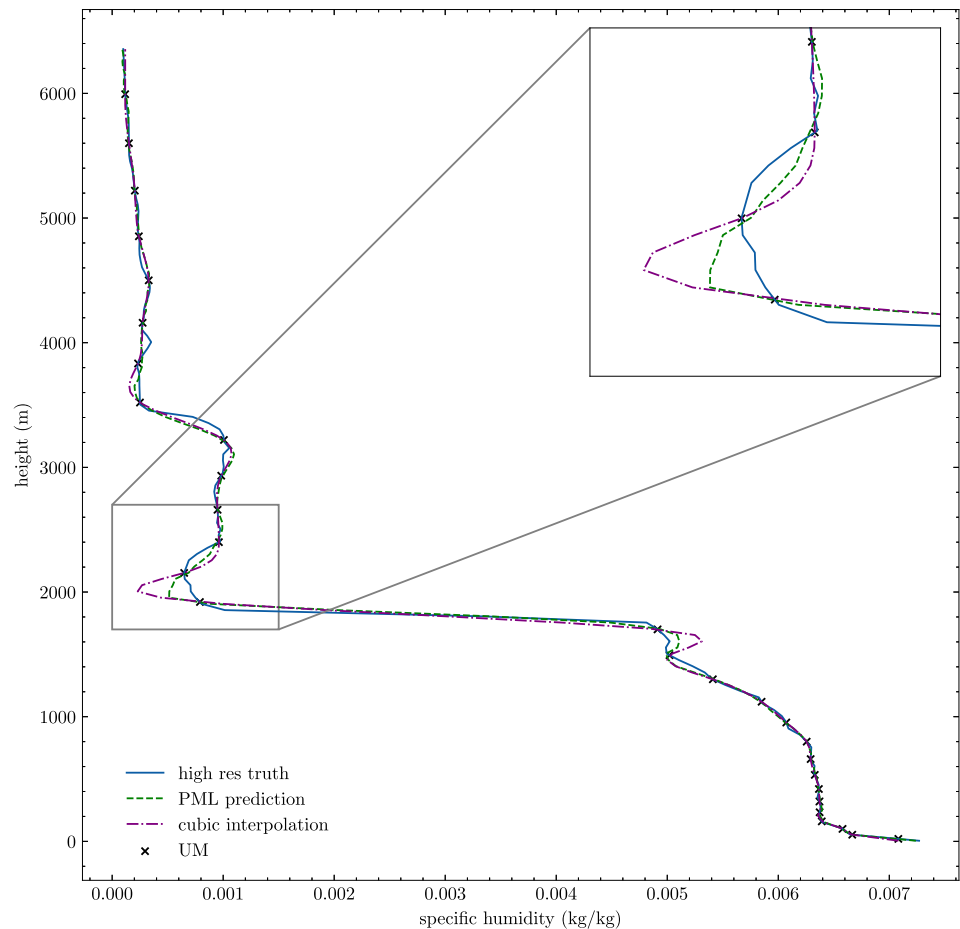
To evaluate if the models' predictions do improve cloud fraction when compared against the original UM resolution data, cloud fraction was calculated from the LR  $T$  and  $q$  profiles and the cloud fraction profiles from the ML models and truth were coarse-grained to LR. When reducing the resolution of the cloud fraction profiles, a 'maximum interpolation' method was used. This is because a simple (linear) interpolation would mean losing the sharp peaks in cloud fraction, which

are what this project aimed to recover in the first place.

To get LR cloud fraction, the maximum value in HR cloud fraction between the middle of the gridbox and the middle of the next gridbox is taken. This can be seen as a representation of cloud area, which is a useful quantity for radiation purposes, for example. The RMSE of the UM profiles compared against LR truth is equal to  $9.335 \times 10^{-2}$  while the traditional ML model gives a RMSE of  $(7.898 \pm 0.554) \times 10^{-2}$  and  $(7.135 \pm 0.112) \times 10^{-2}$  for the physics-informed model.

Refractivity was also investigated. From the different profiles of  $T$  and  $q$  obtained from the truth, cubic

**FIGURE 3** Example of a specific humidity profile from Eastern North Atlantic Graciosa Island, Azores, Portugal (23/11/2014, 22:30:30) predicted by the ML model (dashed green line) compared against the cubic interpolated profile (dash-dotted purple line) and the truth (full blue line). Values on the atmospheric model grid are shown by crosses. The left panel shows the whole profile while the right one focuses on the sudden change in  $q$ .



interpolation and the two versions of the ML model, the refractivity profiles were calculated and compared. Regarding refractivity, the cubic interpolated profiles led to a RMSE value of 1.048, the traditional ML, a RMSE of  $(9.956 \pm 0.012) \times 10^{-1}$  and the PML to  $1.002 \pm 0.002$ . Figure 5 shows a profile of refractivity calculated from profiles of  $T$  and  $q$  predicted by the traditional ML model compared to the truth and the profile obtained from cubic interpolated  $T$  and  $q$ .

Finally, Figure 6 shows three skew- $T$  log- $p$  diagrams obtained from the truth, cubic interpolation and ML predictions (from the traditional ML version) of  $T$  and  $q$ , where it is possible to observe how these different thermodynamic profiles lead to considerably different measurements of CIN, as well as CAPE.

## 4 | DISCUSSION

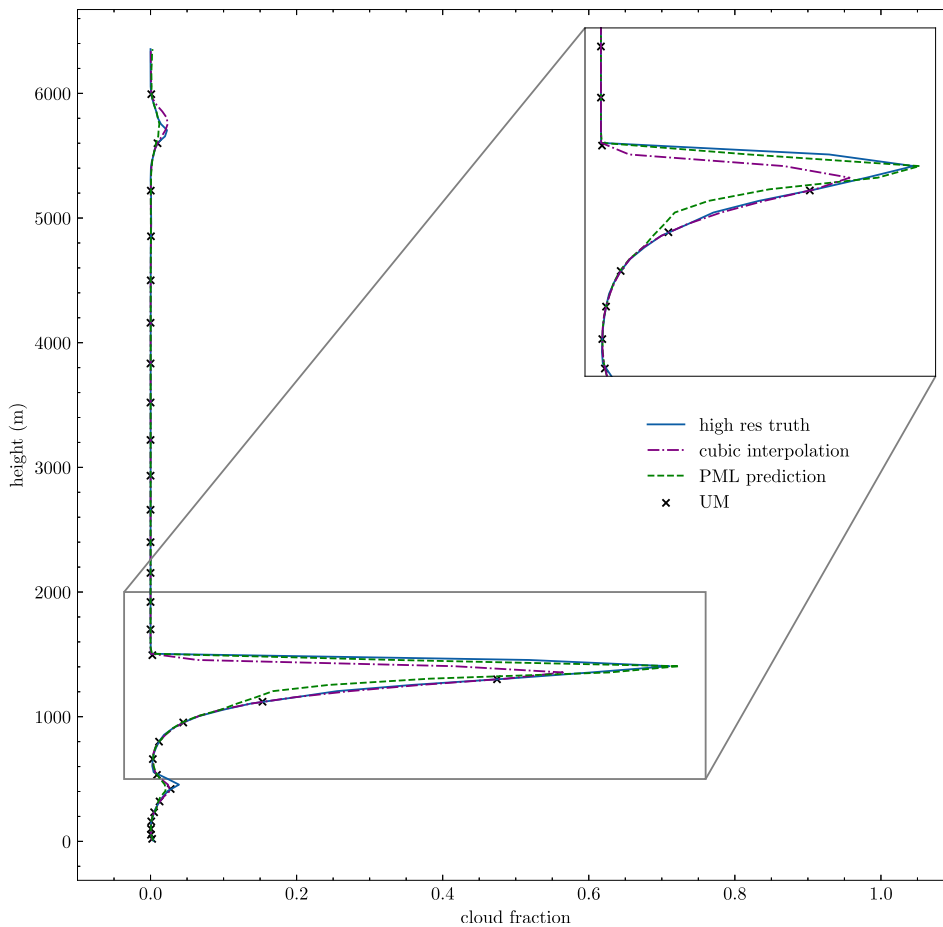
Two versions of a CNN were trained to add vertical detail to atmospheric  $T$  and  $q$  profiles and both models were able to produce refinements that surpass a benchmark provided by cubic interpolation.

The first model just aimed to reduce the error in  $T$  and  $q$ . This model had no knowledge of how other

atmospheric variables are non-linearly related to  $T$  and  $q$ . Although the model did improve the profiles of  $T$  and  $q$ , it did not necessarily improve them at the best places or in the best way to improve cloud fraction.

The second model by contrast had information about how RH is related to  $T$  and  $q$ . This model did lead to cloud fraction profiles which were an improvement above that provided by a simple cubic interpolation of  $T$  and  $q$ . This improvement comes at the cost of the errors in  $T$  and  $q$  themselves being slightly worse, although still better than what a cubic interpolation would provide, as shown by Tables 1 and 2. It is clear from these tables that both ML models improve the profiles of temperature and humidity, in comparison to cubic interpolation, even though the traditional ML model produces most of the times greater benefit than the physics-informed one.

This is further sustained by looking at Table 3, where it can be observed that while the traditional ML can improve RMSE of cloud fraction, when compared against cubic, its physics-informed counterpart does a better job at it, showing an overall improvement in this metric. Although Table 4 shows slightly more mixed results, relative to cloud base height, it does show once again that PML also improves upon tradition ML, for all locations.



**FIGURE 4** Example of cloud fraction (CF) profile from Eastern North Atlantic Graciosa Island, Azores, Portugal (21/11/2014, 05:30:30) obtained from the profiles of  $T$  and  $q$  predicted by the PML model (dashed green line) compared to the one obtained from cubic interpolated profiles (dash-dotted purple line) and the truth (full blue line). Values on the atmospheric model grid are shown by crosses. The left panel shows the profile up to 6 km while the right one zooms in on the sudden peak in CF.

Table 5 shows the results for CIN and it supports what was previously mentioned, that is, that both models perform well and improve overall the RMSE compared against the benchmark. For this variable, the traditional ML version surpasses the physics-informed one. The same is observed for CAPE (Table 6).

Tables 1–6 show that the models do a good job at generalising to a different unseen location and year, with the results for Ascension Island showing an improvement over the cubic interpolation for all variables for at least one of the versions of the ML model.

As well as the quantitative improvements shown in these tables, the examples shown in Figures 2–6 show the algorithm producing other benefits. Figure 2 shows that the algorithm is able to sharpen the temperature inversion better than cubic interpolation while, in terms of humidity, it is shown in Figure 3 that the ML approach sharpens the gradient and does not suffer from the overshoots and undershoots shown by cubic interpolation.

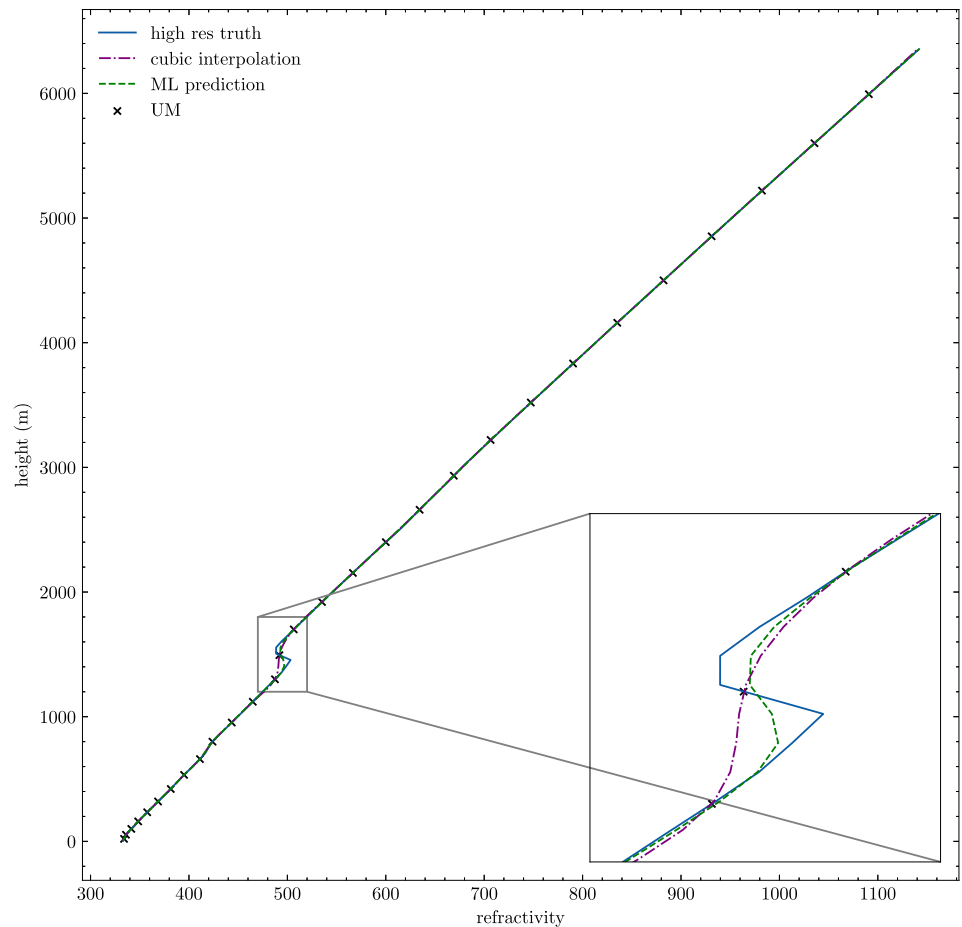
In the cloud fraction profiles shown in Figure 4, as well as an improvement in the lateral extent of the cloud, the height of cloud top is better captured. It is encouraging that the U-Net is also able to capture a region of reversed vertical gradient of refractivity (Figure 5).

Finally, in the example in Figure 6, there is a shallow layer of weak moist stability from 720 to 820 hPa in reality. The cubic interpolation infers that this whole layer is moist neutral, leading to an underestimate in CIN. The U-Net, by contrast, successfully captures a weakly stable moist layer, leading to a better estimate of the amount of CIN, as well as, CAPE.

One approach to improving the representation of fine-scale atmospheric phenomena such as thin clouds, shallow layers of CIN and slim regions of anomalous refractivity gradients is to increase the vertical resolution of the model, for example by doubling the number of layers on which atmospheric variables are discretised. This however would make the model significantly more expensive due to a combination of three factors. First, there would be a doubling of the memory requirements from needing to store all the atmospheric variables on more levels. Second, there would be a doubling of the number of calculations required in all aspects of the model physics and dynamics. Third, the model timestep may need to be reduced to ensure numerical stability.

The new technique presented here however could allow computer resources to be deployed more strategically. Having identified physical processes that are likely

**FIGURE 5** Example of a refractivity profile from Eastern North Atlantic Graciosa Island, Azores, Portugal (07/12/2014, 00:30:30) calculated from the  $T$  and  $q$  profiles predicted by the traditional ML model (dashed green line) compared against the cubic interpolated profile (dash-dotted purple line) and the truth (full blue line). Values on the atmospheric model grid are shown by crosses. The left panel shows the whole profile up to 6 km and the right one zooms in on the sudden inversion.



to benefit from increased vertical resolution, the new CNN could be called just prior to that parametrisation scheme being called. For example, a HR thermodynamics profile could be generated before calling the cloud scheme. The cloud scheme could then be fed this profile and predict cloud fraction and condensate amounts on a HR grid, which could then be coarse-grained back to the model original grid. As a result, thin cloud layers which may otherwise have been missed or under-represented could now appear in the model and be allowed to have a radiative effect. This could lead to significant improvements in model performance, with only a modest increase in computation.

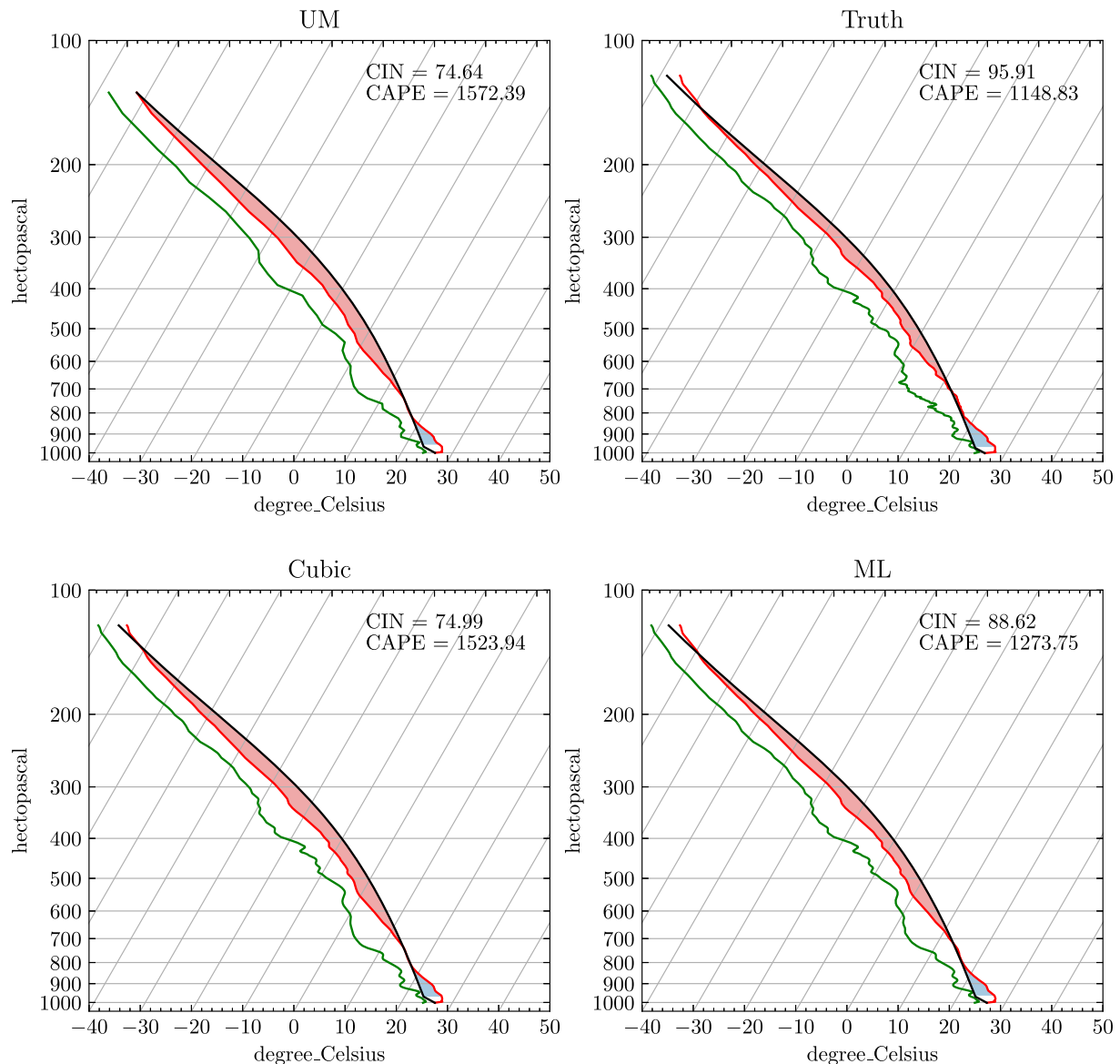
Because of imperfections in existing parts of any atmospheric model, there are compensating errors in the representation of various processes. As a result, improvements to certain aspects of one scheme often require other schemes to be tuned in order to get overall improved performance (Hourdin et al., 2017). Coupling this scheme to the UM would inevitably lead to changes in the statistics of cloud occurrence and convective initiation which would lead to a need to retune several aspects of the model physics. This will complicate the objective assessment of the new approach.

Implementing and evaluating this thermodynamic vertical resolution enhancement technique in the UM is therefore beyond the scope of this study. The focus here has been on the development and evaluation of the technique in offline mode. The technique described here however will be part of a package of multiple parametrisation changes that will form part of a new formal model configuration (e.g. Walters et al., 2017).

## 5 | CONCLUSION

Some phenomena in the atmosphere occur at scales smaller than the grid of our models, particularly in the vertical. An example of this is the formation of thin clouds, which can happen due to sudden temperature inversions in the vertical. While increasing the vertical resolution of atmospheric models would improve the representation of these processes, this is expensive both in memory and computations needed. A possible solution was presented in this paper.

A machine-learning-based one-dimensional SR model was trained to improve the vertical detail in temperature and specific humidity profiles. Radiosonde data were



**FIGURE 6** Example skew- $T$  log- $p$  diagrams from the Tropical Western Pacific Central Facility, Darwin, Australia (06/12/2014, 16:30:30) on the UM vertical grid (top left), from the true high-resolution profiles of  $T$  and  $q$  (top right), from cubic interpolation (bottom left) and from the ML model predictions (bottom right). The shaded red and blue regions represent positive and negative areas, respectively.

used as the truth, due to their high vertical resolution relative to atmospheric models. The first step in the inference process is to take the profiles of  $T$  and  $q$  and regrid to a 50 m grid using cubic interpolation. The CNN then refines the details of the profiles on this grid. This makes this technique general and means it can be used with thermodynamic profiles from any atmospheric model, not just the UM.

Two versions of the ML model were trained, one using a simple MSE as the cost function (traditional ML) and the other using a RH penalty in the cost function (PML). Although the first model performs better than the benchmark (cubic interpolation) at both  $T$  and  $q$ , it does not, in general, improve cloud fraction. This is likely due to the non-linear relations that exist between

temperature, humidity and cloud fraction and might mean that even though the model has learned how to improve the vertical profiles of  $T$  and  $q$ , it is not improving them at the most relevant places to improve cloud fraction.

The PML model, on the other hand, is given information on how  $T$  and  $q$  impact RH, and this leads to an improvement in cloud fraction, which surpasses the benchmark. This improvement in cloud fraction comes at the cost of slightly worse overall profiles of  $T$  and  $q$ , although still better than the cubic interpolated ones.

It was also investigated how CIN, CAPE and refractivity index calculated from the thermodynamic profiles predicted by the ML model compared to the ones calculated from cubic interpolated profiles and HR truth. Both

versions of the model led to improvements in all these three variables, when compared to cubic interpolation.

All experiments in this study were done offline. While offline profile refinement may be sufficient for improving diagnostic calculations of CAPE and CIN and diagnosing regions of anomalous radar refractivity gradients, other applications will require online deployment. As a result, the next step will be to implement this model in the UM, coupled to the cloud scheme, to investigate how it behaves and how it affects the model forecasts.

As shown with CIN, this technique can also be used to improve the representation of other physical processes besides cloud formation. So in the future, it might be coupled to other parametrisations in the UM, such as the convection scheme or any other where a better resolved vertical profile of key variables could lead to better performance.

### AUTHOR CONTRIBUTIONS

**Joana D. da Silva Rodrigues:** Investigation; writing – original draft; writing – review and editing. **Cyril J. Morcrette:** Investigation; writing – original draft; writing – review and editing.

### ACKNOWLEDGEMENTS

Code used in this project is available in <https://github.com/jdsrodrigues/superresolution>. Data were obtained from the Atmospheric Radiation Measurement (ARM) Program sponsored by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Climate and Environmental Sciences Division. Thanks are due to Scott Giangrande from Brookhaven National Laboratory for useful discussions and advice about the radiosonde data used in this project.

### FUNDING INFORMATION

This work was funded by the Met Office Weather and Climate Science for Service Partnership (WCSSP) South-east Asia project, which is supported by the Department for Science, Innovation Technology (DSIT) and by AI4PEX (Artificial Intelligence and Machine Learning for Enhanced Representation of Processes and Extremes in Earth System Models) which has received funding from UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee agreements number 10114295, 10103109 and 10093450.

### CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

### DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Atmospheric Radiation Measurement (ARM) user facility at <http://dx.doi.org/10.5439/1095316>.

### ORCID

Joana D. da Silva Rodrigues  <https://orcid.org/0009-0006-6411-7672>

Cyril J. Morcrette  <https://orcid.org/0000-0002-4240-8472>

### REFERENCES

- Bauer, P., Thorpe, A. & Brunet, G. (2015) The quiet revolution of numerical weather prediction. *Nature*, 525, 47–55.
- Bony, S. & Dufresne, J.-L. (2005) Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophysical Research Letters*, 32, L20806. <https://doi.org/10.1029/2005GL023851>
- Boutle, I.A. & Morcrette, C.J. (2010) Parametrization of area cloud fraction. *Atmospheric Science Letters*, 11, 283–289. Available from: <https://doi.org/10.1002/asl.293>
- 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. *Bulletin of the American Meteorological Society*, 93, 1865–1877.
- Chen, H., Zhang, X., Liu, Y. & Zeng, Q. (2019) Generative adversarial networks capabilities for super-resolution reconstruction of weather radar echo images. *Atmosphere*, 10, 555.
- Fletcher, J.K. & Bretherton, C.S. (2010) Evaluating boundary layer-based mass flux closures using cloud-resolving model simulations of deep convection. *Journal of the Atmospheric Sciences*, 67, 2212–2225.
- Gultepe, I., Sharman, R., Williams, P.D., Zhou, B., Ellrod, G., Minnis, P. et al. (2019) A review of high impact weather for aviation meteorology. *Pure and Applied Geophysics*, 176, 1869–1921. Available from: <https://doi.org/10.1007/s00024-019-02168-6>
- Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q. et al. (2017) The art and science of climate model tuning. *Bulletin of the American Meteorological Society*, 98, 589–602. <https://journals.ametsoc.org/view/journals/bams/98/3/bams-d-15-00135.1.xml>
- Jensen, M., Giangrande, S., Fairless, T. & Zhou, A. (2013) *Interpolated sonde (INTERPOLATEDSONDE). 2013-10-01 to 2014-12-31, eastern North Atlantic (ENA) Graciosa Island, Azores, Portugal (C1), north slope Alaska (NSA) central facility, Barrow AK (C1), southern Great Plains (SGP) central facility, Lamont, OK (C1), tropical Western Pacific (TWP) central facility, Darwin, Australia (C3)*. Data set. Oak Ridge, TN: Atmospheric Radiation Measurement (ARM) User Facility Data Center. Available from: <https://doi.org/10.5439/1095316>
- Kuang, Z. & Bretherton, C.S. (2006) A mass-flux scheme view of a high-resolution simulation of a transition from shallow to deep cumulus convection. *Journal of the Atmospheric Sciences*, 63, 1895–1909.
- Kuleshov, V., Enam, S.Z. & Ermon, S. (2017) Audio super resolution using neural networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Workshop Track Proceedings*. arXiv. Preprint. Available from: <https://doi.org/10.48550/arXiv.1708.00853>
- Manepalli, A., Singh, A., Mudigonda, M., White, B. & Albert, A. (2020) Generalization properties of machine learning based weather model downscaling. In *Proceedings of the International Conference on Learning Representations, ICLR 2020, AI for Earth*

- Sciences Workshop, April 26. Available at: <https://ai4earthscience.github.io/iclr-2020-workshop/papers/ai4earth25.pdf>.
- May, R.M., Goebbert, K.H., Thielen, J.E., Leeman, J.R., Camron, M.D., Bruick, Z. et al. (2022) Metpy: a meteorological python library for data analysis and visualization. *Bulletin of the American Meteorological Society*, 103, E2273–E2284. <https://journals.ametsoc.org/view/journals/bams/103/10/BAMS-D-21-0125.1.xml>
- Mendes, J., Zwane, N., Mabasa, B., Tazvinga, H., Walter, K., Morcrette, C.J. et al. (2024) An analysis of the effects of clouds in high-resolution forecasting of surface shortwave radiation in South Africa. *Journal of Applied Meteorology and Climatology*, 63, 227–244. Available from: <https://doi.org/10.1175/JAMC-D-23-0058.1>
- Odena, A., Dumoulin, V. & Olah, C. (2016) *Deconvolution and checkerboard artifacts*. Distill. <http://distill.pub/2016/deconv-checkerboard/>
- Orhan, A.E. & Pitkow, X. (2017) Skip connections eliminate singularities. In *Proceedings of the International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017*. arXiv. Preprint. Available from: <https://doi.org/10.48550/arXiv.1701.09175>
- Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R. et al. (2016) Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Las Vegas, NV: IEEE, pp. 1874–1883.
- Slingo, A. (1990) Sensitivity of the earth's radiation budget to changes in low clouds. *Nature*, 343, 49–51.
- Smith, R.N.B. (1990) A scheme for predicting layer cloud and their water content in a general circulation model. *Quarterly Journal of the Royal Meteorological Society*, 116, 435–460.
- Stengel, K., Glaws, A., Hettinger, D. & King, R.N. (2020) Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences*, 117, 16805–16815.
- Stensrud, D.J. (2007) *Parameterization schemes: keys to understanding numerical weather prediction models*. Cambridge: Cambridge University Press.
- Stirling, A.J. & Stratton, R.A. (2012) Entrainment processes in the diurnal cycle of deep convection over land. *Quarterly Journal of the Royal Meteorological Society*, 138, 1135–1149. Available from: <https://doi.org/10.1002/qj.1868>
- Van Weverberg, K., Boutle, I.A., Morcrette, C.J. & Newsom, R.K. (2016) Towards retrieving critical relative humidity from ground-based remote-sensing observations. *Quarterly Journal of the Royal Meteorological Society*, 142, 2867–2881. Available from: <https://doi.org/10.1002/qj.2874>
- Walters, D.N., Baran, A., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J. et al. (2017) The met Office unified model global atmosphere 7.0 and JULES global land 7.0 configurations. *Geoscientific Model Development*, 12, 1909–1963. Available from: <https://doi.org/10.5194/gmd-2017-291>
- Wang, C., Wilson, D., Haack, T., Clark, P., Lean, H. & Marshall, R. (2012) Effects of initial and boundary conditions of mesoscale models on simulated atmospheric refractivity. *Journal of Applied Meteorology and Climatology*, 51, 115–132. Available from: <https://doi.org/10.1175/JAMC-D-11-012.1>
- Wang, X., Han, Y., Xue, W., Yang, G. & Zhang, G.J. (2021) Stable climate simulations using a realistic gcm with neural network parameterizations for atmospheric moist physics and radiation processes. *Geoscientific Model Development Discussions*, 2021, 1–35.
- Williams, K.D., Hewitt, A.J. & Bodas-Salcedo, A. (2020) Use of short-range forecasts to evaluate fast physics processes relevant for climate sensitivity. *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001986. Available from: <https://doi.org/10.1029/2019MS001986>
- Wood, R. & Field, P.R. (2000) Relationships between total water, condensed water, and cloud fraction in stratiform clouds examined using aircraft data. *Journal of the Atmospheric Sciences*, 57, 1888–1905.
- Wu, Y., Teufel, B., Sushama, L., Belair, S. & Sun, L. (2021) Deep learning-based super-resolution climate simulator-emulator framework for urban heat studies. *Geophysical Research Letters*, 48, e2021GL094737.

**How to cite this article:** da Silva Rodrigues, J. D., & Morcrette, C. J. (2025). Improving vertical detail in simulated temperature and humidity data using machine learning. *Atmospheric Science Letters*, 26(2), e1288. <https://doi.org/10.1002/asl.1288>