



# Representing subgrid-scale cloud effects in a radiation parameterization using machine learning: MLe-radiation v1.0

Katharina Hafner<sup>1,2</sup>, Sara Shamekh<sup>3</sup>, Guillaume Bertoli<sup>4</sup>, Axel Lauer<sup>2</sup>, Robert Pincus<sup>5</sup>, Julien Savre<sup>2</sup>, and Veronika Eyring<sup>2,1</sup>

<sup>1</sup>University of Bremen, Institute of Environmental Physics (IUP), Bremen, Germany

<sup>2</sup>Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

<sup>3</sup>Courant Institute of Mathematical Sciences, New York University (NYU), New York, NY, USA

<sup>4</sup>Department of Earth and Environmental Engineering, Columbia University, New York, NY, USA

<sup>5</sup>Lamont-Doherty Earth Observatory, Palisades, New York, USA

**Correspondence:** Katharina Hafner (hafner@iup.physik.uni-bremen.de)

Received: 6 October 2025 – Discussion started: 16 October 2025

Revised: 6 March 2026 – Accepted: 7 April 2026 – Published: 13 May 2026

**Abstract.** Improvements of Machine Learning (ML)-based radiation emulators remain constrained by the underlying assumptions to represent horizontal and vertical subgrid-scale cloud distributions, which continue to introduce substantial uncertainties. In this study, we introduce a method to represent the impact of subgrid-scale clouds by applying ML to learn processes from high-resolution model output with a horizontal grid spacing of 5 km. In global storm resolving models, clouds begin to be explicitly resolved. Coarse-graining these high-resolution simulations to the resolution of coarser Earth System Models yields radiative heating rates that implicitly include subgrid-scale cloud effects, without assumptions about their horizontal or vertical distributions. We define the cloud radiative impact as the difference between all-sky and clear-sky radiative fluxes, and train the ML component solely on this cloud-induced contribution to heating rates. The clear-sky tendencies remain being computed with a conventional physics-based radiation scheme. This hybrid design enhances generalization, since the machine-learned part addresses only subgrid-scale cloud effects, while the clear-sky component remains responsive to changes in greenhouse gas or aerosol concentrations. Applied to coarse-grained data offline, the ML-enhanced radiation scheme reduces errors by a factor of 4–10 compared with a conventional coarse-scale radiation scheme. This shows the potential of representing subgrid-scale cloud effects in radiation schemes with ML for the next generation of Earth System Models.

## 1 Introduction

For climate projections, coarse-scale Earth System Models (ESMs) typically have horizontal resolutions of 100–200 km (Chen et al., 2021), in which clouds cannot be resolved explicitly. Therefore, these models require parameterizations of fractional cloudiness, particularly for cloud–radiation interactions at subgrid scales. A widely used approach is the Monte Carlo Independent Column Approach (McICA) (Pincus et al., 2003), where g-points are randomly assigned as cloudy or clear-sky according to the cloud fraction. This stochastic simplification introduces uncertainties in cloud–radiation interactions of up to  $100 \text{ W m}^{-2}$  in surface fluxes, corresponding to relative errors of 10 % or more. However, these errors are unbiased compared to the Independent Column Approach (ICA) (Barker et al., 1999), where entire subcolumns are randomly designed as cloudy or clear, and thus tend to average out in long ESM simulations (Pincus et al., 2003). Vertical cloud overlap is commonly parameterized using the maximum-random overlap assumption (Räisänen et al., 2004), whereby adjacent layers overlap maximally while distant cloud layers overlap randomly. Alternatively, some models employ an *all-or-nothing* cloud cover scheme, which is a good approximation in high-resolution simulations (Giorgetta et al., 2022; Hohenegger et al., 2023).

Machine learning (ML)-based radiation emulators have been developed for more than two decades (Ukkonen, 2022; Yao et al., 2023; Roh and Song, 2020; Pal et al., 2019;

Krasnopolsky et al., 2005; Lagerquist et al., 2023). One potentially appealing aspect of ML-based emulators is their relative speed compared to traditional radiative transfer models, which in principle allows more frequent calls during ESM integration. In practice, however, the speed-up potential has proven limited (Bertoli et al., 2025; Ukkonen and Chantry, 2025; Hafner et al., 2025a), with faster performance achievable through code optimization (Cotronei and Slawig, 2020; Ukkonen and Hogan, 2024). Moreover, improvements from more frequent radiation calls (Hafner et al., 2025a) remain marginal. Nonetheless, ML-based radiation emulators are valuable in specific context, such as differentiable ESMs (Kochkov et al., 2024; Klöwer et al., 2024) – where the derivative can be calculated for a complete model integration step – or GPU-based ESMs, where they are more energy efficient than physics-based schemes (Ukkonen and Chantry, 2025). Despite their high accuracy, emulators still face challenges in cloudy conditions (Hafner et al., 2025b), which may partly explain why they have not yet shown substantial improvements over traditional radiation schemes.

As more high-resolution Global Storm Resolving Model (GSRM) data become available, they offer increasing opportunities to enhance coarse-scale ESMs with machine learning. High-resolution simulations have been applied to nudge coarse-scale models toward fine-scale states (Bretherton et al., 2022), to learn subgrid tendencies directly (Heuer et al., 2024; Busecke et al., 2025), to infer subgrid effects from coarse-scale states (Grundner et al., 2022; Shamekh et al., 2023) and to learn all physics parameterizations (Watt-Meyer et al., 2024). Beyond these applications, high-resolution model simulations enable new strategies for representing fractional cloudiness and cloud overlap in coarse-scale ESMs using ML. Specifically, GSRM output can be coarse-grained to the resolution of the target ESM, and a neural network (NN) can then be trained to learn the subgrid-scale distribution of clouds from the underlying statistics. For instance, Henn et al. (2024) predicted coarse-grained cloud fields to reduce radiative biases; however, the cloud overlap assumption in the radiation scheme remains unchanged, limiting improvements. Leveraging ML to represent subgrid-scale cloud effects on radiation thus provide a promising pathway toward accurate radiation schemes in ESMs for climate projections.

In this work, we demonstrate how the representation of cloud radiative impacts on heating rates can be improved by separating them from the all-sky heating rates and learning the cloud contribution directly from high-resolution simulations. Because these simulations explicitly resolve cloud systems without relying on assumptions about their horizontal and vertical distributions, the ML model can implicitly account for subgrid-scale cloud effects on the radiative heating rates. A similar separation between clear-sky and cloudy radiative fluxes was previously proposed by Chevallier et al. (1998). Moreover, Meyer et al. (2022) focused on learning 3D cloud radiative effects. However, the explicit separation

of cloud radiative impacts from all-sky radiation, especially including subgrid-scale effects, remains largely unexplored.

We aim to use ML to encode the true variability of vertical and horizontal subgrid-scale clouds and their radiative impacts from high-resolution simulations, rather than relying on statistical schemes. This raises the central question: “Can the representation of cloud–radiation interactions be improved by training on high-resolution simulation data?” To address this, we compare our ML approach against a state-of-the-art physics-based radiative transfer model, namely RTE+RRTMGP (Pincus et al., 2019) with McICA (Pincus et al., 2003) and maximum-random overlap (Räisänen et al., 2004) to represent subgrid-scale cloudiness. In particular, we evaluate how a physics-based coarse-scale radiation scheme performs when applied to coarse-grained data. This comparison allows us to disentangle differences arising from changes in the input parameters from those due to different resolutions.

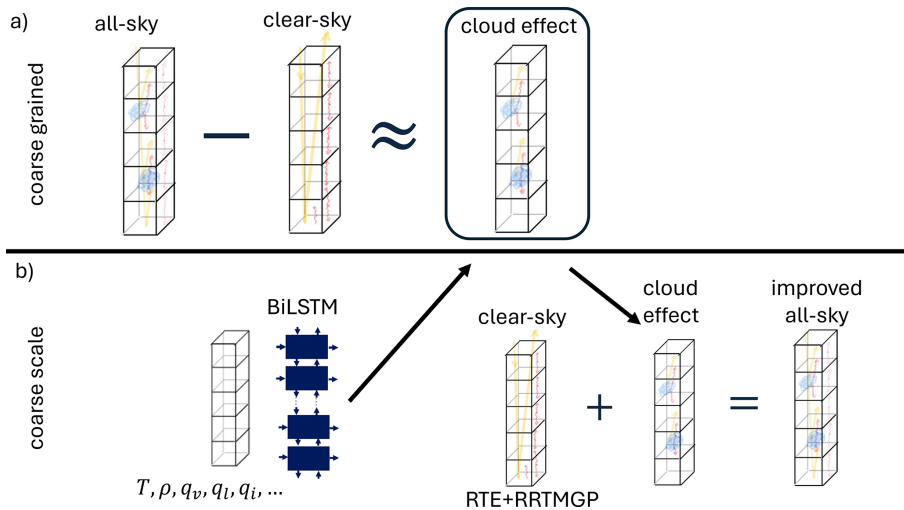
This paper is structured as follows. Section 2 describes the method used to learn the cloud radiative impact from high-resolution simulation data. Section 3 introduces the main datasets, generated with the ICOSahedral Non-hydrostatic (ICON) model (Giorgetta et al., 2018, 2022) in both high- and coarse-resolution configurations, and includes a comparison of the input and output variables of the radiation parameterization across these resolutions. In Sect. 4, we present the main results by comparing the physics-based coarse-scale radiation parameterization with the ML-enhanced scheme on coarse-grained test data. Finally, Sect. 5 provides a discussion and conclusion.

## 2 Learning the Cloud Radiative Impact on Heating Rates

We define the cloud radiative impact on heating rates in a column as the residual between the all-sky and clear-sky heating rates, where clear-sky represents the same atmospheric conditions as all-sky but without clouds (Fig. 1a). In ESMs such as the ICON model (Giorgetta et al., 2018, 2022), radiation parameterizations like RTE+RRTMGP (Pincus et al., 2019) first calculate gas optical properties and then add cloud optical properties. Then, these combined properties are used to calculate all-sky radiative fluxes, and clear-sky fluxes can be obtained omitting the cloud optical properties. Heating rates are then derived from the flux divergence for both all-sky and clear-sky conditions. The residual of these heating rates, which is the cloud radiative impact (CRI) on the heating rates, serves as the training target for the ML-based radiation parameterization:

$$\frac{\partial T_{\text{CRI}}}{\partial t} = \frac{\partial T_{\text{all-sky}}}{\partial t} - \frac{\partial T_{\text{clear-sky}}}{\partial t}. \quad (1)$$

The heating rates  $\frac{\partial T_k}{\partial t}$  in a layer  $k$  are calculated from the net flux  $F_{\text{Net}}$  at the layer boundaries  $k \pm \frac{1}{2}$



**Figure 1.** Sketch of constructing the cloud radiative impact on heating rates. Radiation schemes calculate fluxes for the same scene once with and once without clouds resulting in all-sky and clear-sky fluxes. The corresponding heating rates can be inferred from the fluxes and the residual yields an approximation of the cloud radiative impact on heating rates for all layers in a column.

$$\frac{\partial T_k}{\partial t} = \frac{F_{\text{Net},k+1/2} - F_{\text{Net},k-1/2}}{c_v m_{\text{air}}}, \quad (2)$$

where  $m_{\text{air}}$  is the mass of moist air per area, and specific heat at constant volume  $c_v$  scales with the tracer mixing ratios.

The central idea is to train a neural network (NN) that learns only the cloud impact on heating rates (see Fig. 1). Cloud–radiation interactions are subject to large uncertainties, since the subgrid-scale horizontal and vertical distributions of clouds are not resolved in coarse-scale ESMs. In the hybrid ML-physics radiation parameterization (Fig. 1b), the NN predicts only the cloud radiative impact, while the clear-sky component is still computed by the original radiation parameterization. This design ensures that the ML-enhanced radiation scheme retains sensitivity to changes in GHG and aerosols through the clear-sky part, potentially improving generalization across different climates. The ML-cloud component can respond to GHG and aerosols changes only indirectly, for example through modifications of the cloud distribution. However, this hybrid approach does not capture secondary effects arising from reflected radiation. The validity is further discussed in the Appendix D.

At first glance, a linear decomposition of the clear-sky heating rate and the cloud radiative impacts on heating rates may seem counterintuitive, given the inherently nonlinear interaction among specific humidity, clouds, and trace gases such as ozone (Bony et al., 2015). Nevertheless, we adopt a linear decomposition framework in this work, as illustrated on Fig. 1. Within this framework, the NN is tasked with learning the nonlinear relationships based on the prevailing atmospheric conditions and cloud-related variables (e.g., cloud liquid water and cloud ice).

## 2.1 Method

We use a bidirectional long short term memory (BiLSTM) based on Hafner et al. (2025b) to learn the cloud radiative impact in an atmospheric column. Bidirectional architectures are particularly well-suited for radiative transfer problems (Ukkonen, 2022; Yao et al., 2023; Ukkonen and Chantry, 2025; Bertoli et al., 2025). Unlike their common usage for temporal sequences, BiLSTMs for radiation scan the vertical dimension in both directions, resembling upward and downward fluxes. The NN consists of one BiLSTM layer with  $\tanh$  activation and a hidden dimension of 96, with two LSTMs scanning the input vertical profiles in upward and downward directions. The combined output of the LSTMs is then processed by a dense layer that predicts the heating rate at each level. The training is split between shortwave (SW) and longwave (LW) radiation as shortwave temperature tendencies are only calculated for sun-lit areas and longwave temperature tendencies are always computed, totaling in 82k trainable parameters per NN.

The input variables are vertical profiles of mass mixing ratios of specific humidity  $q_v$ , cloud liquid  $q_l$ , cloud ice  $q_i$ , snow  $q_s$  and ozone  $\text{O}_3$ , plus the cloud fraction  $cl$ , air density  $\rho$ , and temperature  $T$ . For SW, we additionally use surface albedo  $\alpha$  and incoming solar flux at the top of the atmosphere  $F_{\downarrow, \text{TOA}, \text{SW}}$ , which is the solar constant weighted by the solar zenith angle and change in Earth–Sun distance to account for daily and seasonal variations. For LW, we additionally use the surface temperature  $T_{\text{surf}}$  as input. As mentioned above, the output is the cloud radiative effect on heating rates derived from all-sky and clear-sky heating rates (Fig. 1).

Normalization is important for faster convergence of the training (LeCun et al., 2012) and generalization (Beucler

et al., 2024).  $O_3$ ,  $\rho$ ,  $T$ , and  $T_{\text{surf}}$  are normalized using their mean values  $\mu$  and standard deviation  $\sigma$ , also known as  $z$ -score normalization ( $x_{\text{norm}} = \frac{x-\mu}{\sigma}$ ). The normalization factors are computed using all cells of one time step and are one-dimensional such that the vertical structure of the variables remain. The vertical structure is an important aspect that the BiLSTM uses to make a vertically correlated prediction. The water related variables  $q_1$ ,  $q_i$ ,  $q_s$  are normalized by the ambient total (radiatively active) water ( $q_v + q_l + q_i + q_s$ ). Here, radiatively active means used in the radiation scheme.  $q_v$  uses  $z$ -score normalization providing information that relates to the absolute mass mixing ratios.  $cl$  and  $\alpha$  are not normalized as they already vary between 0 and 1.  $F_{\downarrow, \text{TOA, SW}}$  is normalized by  $1360 \text{ W m}^{-2}$ , which is close to the solar constant. The cloud radiative impact on heating rates is only converted to  $\text{K d}^{-1}$ , which is on the order of one.

The loss we minimize during training consists of the sum of mean absolute error (MAE) and mean squared error (MSE). We use the Adam optimizer (Kingma and Ba, 2017), and set a learning rate of  $10^{-3}$ , which is reduced by a factor of 2 when the validation loss does not decrease by 0.1 % for 5 epochs. To avoid overfitting, we use early stopping, which stops the training if the validation loss does not decrease for 10 epochs.

### 3 Data

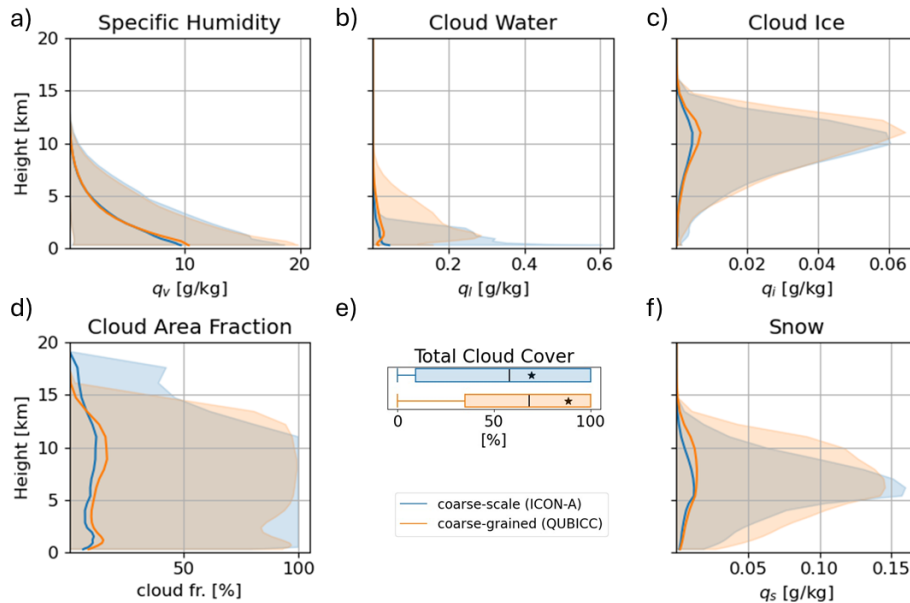
ICON is a weather and climate model permitting simulations across different resolutions, from a few to hundreds of kilometers. For global and long-term applications, the atmospheric component ICON-A (Giorgetta et al., 2018) often runs at 80 km horizontal resolution, and 47 vertical levels covering the altitude range 0–83 km, with parameterizations for radiation, cloud microphysics, turbulence, convection and gravity waves. The ICON model has the option to be run as a GSRM (Stevens et al., 2019; Giorgetta et al., 2022; Hohenegger et al., 2023). The high-resolution simulations used here follow the QUBICC (Quasi-Biennial oscillation in a changing climate) protocol from Giorgetta et al. (2022). The QUBICC simulations have 5 km horizontal resolution, and the vertical dimension spans 83 km on 191 levels. The high-resolution allows the model to run without a convection scheme and gravity wave parameterization, as these processes are starting to be resolved.

We performed QUBICC simulations that cover a total of 40 d evenly distributed across four months: November 2004, January, April, July 2005. The simulations have a physics time step of 40 s and a radiation time step of 6 min. These simulations are run with prescribed sea surface temperatures, sea ice concentrations, greenhouse gas concentrations but no aerosols. The outputs are saved every 192 min. The uneven output interval was chosen to cover a large variability of different solar zenith angles at different locations. The first 6 d of each 10 d period is used for training, the

next 2 d for validation and the last 2 d for testing. Using QUBICC data to train our ML model for ICON-A requires coarse-graining the high-resolution QUBICC dataset. All variables are horizontally and vertically coarse-grained from high-resolution simulations as in Grundner et al. (2022). For horizontal remapping, the high-resolution cells are weighted by their cell area that is contained in each coarse cell. High-resolution cells can be contained fully or partially in a coarse cell, which is represented by a smaller cell area. Similarly, the layer thickness was used for vertical coarse-graining, corresponding to a weighted average. This method is valid for all input and output variables used here, i.e., concentrations and fluxes. Cloud-related variables are expressed as mass fraction, ensuring consistent vertical coarse-graining. Absolute mass variables, such as air mass  $m_{\text{air}}$  in Eq. (2), are coarse-grained by the (weighted) vertical sum that are partially or fully contained in a coarse layer. We discarded a small number of coarse-grained cells if their surface height showed inconsistencies, which can occur over complex terrain. Specifically, the coarse-grained surface height was computed from the high-resolution grid file and compared to the surface height from the coarse-scale grid file, which is slightly rotated and shifted. Consequently, some high-resolution cells are only partially contained in a coarse cell, which can lead to a small mismatch in surface height. Cells with deviations of more than 0.5 m in surface height were discarded. Then, we randomly sampled 35k and 5k grid points per time step for the training and validation set. For the test set, we randomly selected 75k cells per time step for LW and 35k cells for SW. This yields 2.3 million training samples, 260k validation samples, 1.9 million test samples for shortwave and 4.2 million test samples for longwave radiation.

For comparison with the high-resolution data, we also made coarse-scale ICON-A simulations for the same periods and with a similar configuration at 80 km horizontal resolution to compare differences in various distributions. The ICON-A simulations are based on the version 2.6.4 described in Giorgetta et al. (2018), while the QUBICC simulations are based on the version icon-2024.10 (ICON Partnership et al., 2024). To make the coarse-scale simulation more comparable, we used the same microphysics scheme (Doms et al., 2011) and no aerosols. See Appendix for a comparison with the default microphysics scheme. Both simulations use the radiation scheme RTE+RRTMGP (Pincus et al., 2019). ICON-A uses McICA and maximum-random overlap to represent subgrid-scale cloud-radiative impacts. In the QUBICC simulation, the *all-or-nothing* scheme is used, assuming horizontal homogeneous distribution of clouds. The physics and radiation time step is 6 min for the ICON-A simulation, matching the radiation time step in QUBICC.

The remaining differences between the coarse-scale ICON-A and high-resolution QUBICC simulation include the horizontal and vertical resolution, the higher temporal resolution of physical process in QUBICC, which is required due to the high spatial resolution. Moreover, QUBICC in-



**Figure 2.** Distributions of water related input variables for ICON-A and coarse-grained QUBICC data. The bold line shows the mean and the shaded area shows 95 % of the spread between the 2.5 % percentile and the 97.5 % percentile. The boxplot is limited by the minimum and maximum values. The box edges are defined at the 25 % and 75 % percentile of the distribution. The black line illustrates the mean of the distribution and the star is the median.

tends to resolve gravity waves and (deep) convection. Additionally, the radiation scheme RTE+RRTMGP (Pincus et al., 2019) in QUBICC uses snow mixing ratio as an input. Other differences could be due to differences between the code versions and tuning, as we did not retune ICON-A.

### 3.1 Comparison of input and output variables

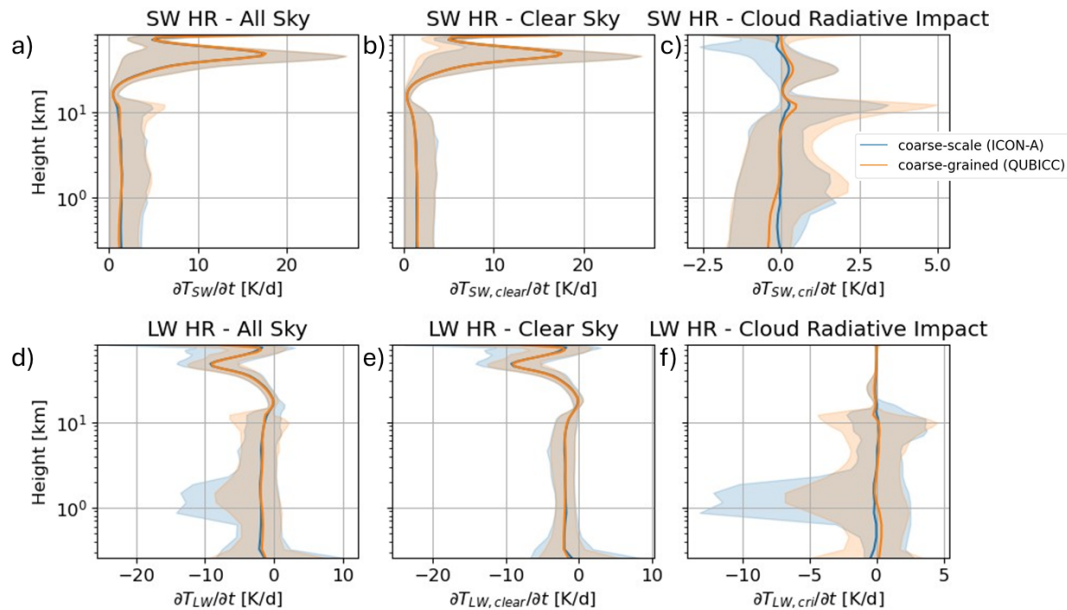
If an ML-based scheme is trained on high-resolution simulations like QUBICC and the goal is to couple it to a coarse-scale model like ICON-A, one needs to ensure that the distributions of input and output variables match. Otherwise, the ML-based scheme could be faced with out-of-distribution samples, which can lead to errors that quickly build-up while the model is integrated (Rasp, 2020). Therefore, we analyze systematic differences between the coarse-scale ICON-A simulations and the high-resolution QUBICC simulations. This analysis is conducted by comparing the spatial and temporal means and spread in the input and output used and produced by the radiation parameterization. Specifically, we focus on  $q_v$ ,  $q_l$ ,  $q_i$ ,  $cl$ , total cloud cover,  $q_s$ , as well as long-wave and shortwave heating rates. For the comparison, we use all samples in the test set. The samples of both simulations cover the same time period which is November 2004, January, April, and July 2005. Comparing two simulations with different grids enables us to uncover systematic differences, with a focus on identifying larger variations.

Distributions of water related input variables are shown in Fig. 2. The distributions of specific humidity look similar in the coarse-grained and coarse-scale data set, where the max-

imum differences between the means is  $0.9 \text{ g kg}^{-1}$  (Fig. 2a). The spread in humidity values also overlaps for both simulations. Cloud water has higher values below 3 km for the coarse-scale simulation while cloud water is more evenly distributed throughout the troposphere for the QUBICC simulations (Fig. 2b). Here, the maximum difference of the mean values is  $0.02 \text{ g kg}^{-1}$ . The distributions of cloud ice have similar shapes, but the coarse-grained distribution peaks higher by  $0.002 \text{ g kg}^{-1}$  (Fig. 2c). The mean cloud area fraction is on average larger by 3 % for the coarse-grained data set below 14 km (Fig. 2d). Snow peaks between 5–10 km (Fig. 2f). However, the vertical distribution is wider in the coarse-grained simulation. Nevertheless, snow was not used as input for radiation in ICON-A.

For comparability, the coarse-grained heating rates from QUBICC had to be rescaled by a factor  $c_v/c_p$  to account for differences in the two model versions used, where  $c_p$  is specific heat at constant pressure.

For SW all-sky heating rates, the mean profiles match very well and the mean difference is  $0.18 \text{ K d}^{-1}$  (Fig. 3a). There are only small differences in the spread of heating rates in the troposphere. The heating rate is decomposed into a clear-sky heating rate – which is calculated from the clear-sky fluxes – and the cloud radiative impact on heating rates. Their distributions are also shown in Fig. 3b and c. The SW clear-sky heating rate has a mean difference of  $0.11 \text{ K d}^{-1}$  for the mean profiles. For the SW cloud radiative impact, the mean difference is also  $0.11 \text{ K d}^{-1}$ . Here, we expect small differences due to mostly resolved clouds in the coarse-grained



**Figure 3.** The distribution of shortwave (top row) and longwave (bottom row) heating rates in coarse-scale and scaled coarse-grained data. The bold line shows the mean and the shaded area shows 95 % of the spread, which is defined as the spread between the 2.5 % percentile and the 97.5 % percentile. The left column shows all-sky heating rates as it is used in the ICON model. The middle column shows clear-sky heating rate computed from clear-sky fluxes, which is a diagnostic output in the ICON model. The right column shows the cloud radiative impact on the heating rate which was computed by subtracting the clear-sky heating rate from the all-sky heating rate.

dataset. However, there is no clear bias between coarse-scale and coarse-grained cloud impacts.

For the LW all-sky heating rates, the mean profiles match well and have a mean difference of  $0.18 \text{ K d}^{-1}$  (Fig. 3d). However, the spread in heating rates is slightly different, which co-occurs with the different spread in cloud water at 1 km (Fig. 2b). The LW clear-sky heating rates look very similar in their mean values and their spread where the mean difference is  $0.14 \text{ K d}^{-1}$  (Fig. 3e). The LW cloud impact is concentrated in the troposphere (Fig. 3f). The mean impact is almost the same between coarse-scale and coarse-grained simulations with a mean difference of  $0.09 \text{ K d}^{-1}$ . However, there is a difference in spread, which can be expected, because the coarse-grained data set implicitly includes subgrid-scale cloud effects.

In the unscaled comparison, the clear-sky heating rates show the same bias as the all-sky heating rates for both SW and LW (Fig. B1). However, this bias does not directly translate to the cloud radiative impact on heating rates because adding the cloud impact is a highly non-linear process (Bony et al., 2015). Additionally, this indicates that the mean cloud effect is similar for a (quasi)-hydrostatic and a non-hydrostatic model.

## 4 Results

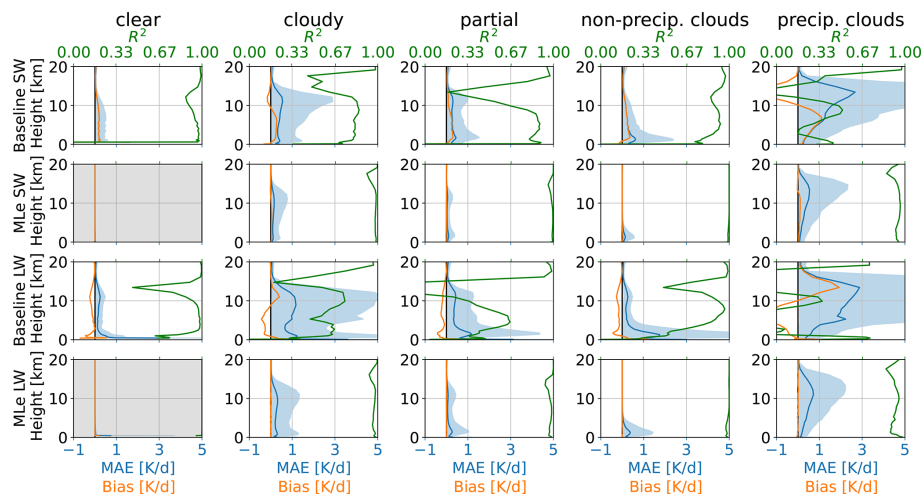
As mentioned above, the simulation outputs of QUBICC and ICON-A can't be used for a sample-by-sample comparison.

Therefore, a coarse-scale radiation reference is needed that is computed (offline) based on coarse grained QUBICC output. All the dataset used in this study are summarized in Table 1.

For the coarse-scale radiation reference, we use the Python front-end (hereafter pyRTE) (Pincus et al., 2025) of the radiation scheme RTE+RRTMGP (Pincus et al., 2019), which is used in all our simulations. Using pyRTE, we replicate the implementation of radiation in QUBICC as closely as possible. For subgrid-scale variability, we employ McICA (Pincus et al., 2003) together with maximum-random cloud overlap (Räisänen et al., 2004). The procedure is as follows: first we calculate gas optical properties, assign them to the atmospheric state and calculate clear-sky fluxes. Next, we compute cloud optical properties, apply McICA with maximum-random overlap to represent subgrid-scale variability, and add cloud optical properties to the gas optical properties. Because QUBICC also includes snow in its radiation parameterization, we additionally calculate snow optical properties and combine them with the other optical properties, before computing all-sky fluxes. Then, heating rates are obtained applying Eq. (2). We calculate heating rates for all samples in the test dataset. The output of pyRTE is used as baseline as it represents the coarse-scale radiation scheme. This allows a sample-by-sample comparison to the coarse-grained heating rates obtained from QUBICC (first and third row in Figs. 4 and 5). The ML-enhanced radiation scheme predicts the cloud radiative impact, which is added to clear sky heating. The resulting all-sky heating is compared to the

**Table 1.** All datasets used in this study. Online means that a simulation was run in high or coarse resolution and the output was saved for certain time steps. Offline means that a parameterization (RTE+RRTMGP or MLe-radiation) computed its tendencies based on model output.

Name	Offline/Online	Purpose
QUBICC	Online	High-resolution simulation that is coarse-grained to serve as training set, compared to ICON-A in Sect. 3 and “ground truth” in Sect. 4.
ICON-A	Online	Coarse-scale simulation used in Sect. 3 to identify systematic biases between coarse-scale and high-resolution simulations.
Baseline	Offline	Calculated using pyRTE based on coarse-grained samples from QUBICC to provide the coarse-scale reference of radiative tendencies, serves as “baseline” and used to measure improvement.
MLe-radiation	Offline	New ML-enhanced radiation parameterization trained on coarse-grained samples from QUBICC computes tendencies, the tendencies are compared to coarse-grained QUBICC and baseline tendencies.

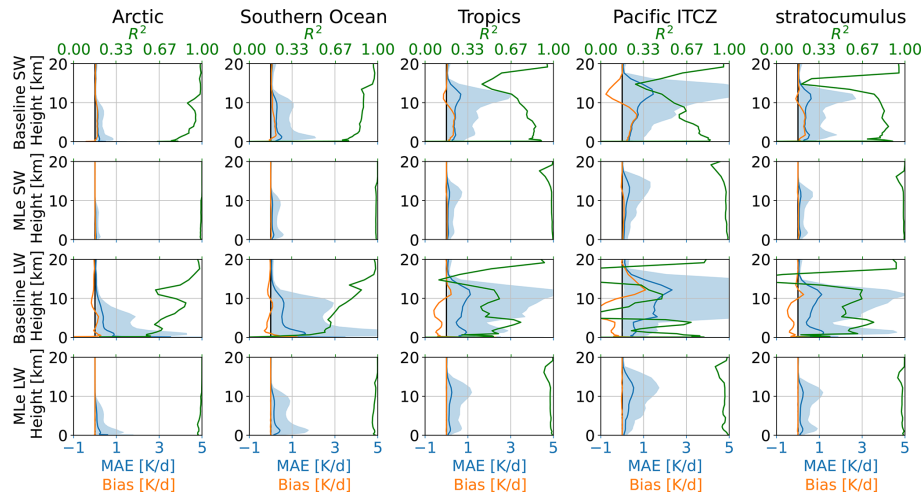


**Figure 4.** Comparison of the baseline and the ML-enhanced radiation scheme on coarse-grained QUBICC data. Results are shown for the shortwave (top rows) and longwave (bottom rows) spectral range. The results are shown separately for clear-sky samples (no clouds, left column), fully cloudy sky samples (second column), and samples with partial cloudiness (third column). Additionally, the results are separated by non-precipitating (fourth column) and precipitating clouds (fifth column). The shown metrics are coefficient of determination  $R^2$  (green), bias (orange) and MAE (blue) with 95 % of the spread, which is defined as the spread between the 2.5 % percentile and the 97.5 % percentiles. The bias and MAE share the  $x$ -axis. The ML-clear-sky panels are gray because the clear-sky fluxes are not calculated by the ML model, but shown as reference.

coarse-grained QUBICC heating rates (second and third row in Figs. 4 and 5). As evaluation metric, we use mean absolute error (MAE), bias, and the coefficient of determination ( $R^2$ ). The improvement is measured by comparing MLe-radiation to the baseline. The results are presented in Fig. 4 with the column-averaged metrics summarized in Table A1. We restrict the presentation to the lowest 20 km of the atmosphere, as cloud impacts are most relevant in the troposphere (see Fig. 3). Nevertheless, the NN predicts the cloud impact on the entire atmospheric column, and the results for the full column can be found in the Appendix C1.

The first column of Fig. 4 shows clear-sky heating rates, for which the cloud distribution plays no role. Accordingly, only small and statistically insignificant differences are ex-

pected, arising mainly from variability in water vapor. While coarse-grained QUBICC heating rates reflect subgrid-scale variability of water vapor, coarse-scale radiation schemes assume a horizontal homogeneous distribution of water vapor. Therefore, the assumption of horizontally homogeneous input parameters introduces a small error of  $0.367 \text{ K d}^{-1}$  (SW) and  $0.571 \text{ K d}^{-1}$  (LW) for the baseline. For comparison with other studies, Table A1 also reports the root mean squared error (RMSE). For clear-sky heating rates of the coarse-scale baseline, the RMSE is  $0.443 \text{ K d}^{-1}$  (SW) and  $0.688 \text{ K d}^{-1}$  (LW) compared with the coarse-grained QUBICC heating rates. Hogan and Matricardi (2022) developed a fast tool for computing gas-optical properties and reports an RMSE of  $0.18 \text{ K d}^{-1}$ . Although the error source differs (gas optical



**Figure 5.** As Fig. 4, but for the selected regions shown in Fig. 5 of Bock and Lauer (2024).

properties vs. spatial resolution), the magnitudes are comparable.

For completeness, we also evaluated the ML-model on clear-sky samples. However, it is not intended for clear-sky scenes as the cloud impact is zero. Therefore, the corresponding panels are grayed out. The MAE is  $0.049 \text{ K d}^{-1}$  for SW and  $0.028 \text{ K d}^{-1}$  for LW.

The second column of Fig. 4 shows results for fully cloudy samples (total cloud cover of 100 %). For the baseline, the MAE peaks near 10 km, exceeding  $0.5 \text{ K d}^{-1}$  for SW and  $1 \text{ K d}^{-1}$  for LW. The corresponding  $R^2$  are low, with averaged values of 0.83 (SW) and 0.66 (LW), compared to 0.98 for the ML-enhanced scheme. The average  $R^2$  values are computed by averaging over the vertical levels.  $R^2$  is weighted by variability, and an  $R^2$  of zero indicates that the error is as large as the variability itself. Since, McICA within a coarse-scale radiation parameterization is supposed to produce unbiased noise, the  $R^2$  is therefore a less informative metric. In contrast, the bias reveals that the coarse-scale radiation scheme systematically struggles to represent the cloud impact near 10 km, particularly for SW. The ML-enhanced scheme produces nearly unbiased heating rates, with MAEs of  $0.106 \text{ K d}^{-1}$  (SW) and  $0.127 \text{ K d}^{-1}$  (LW), representing errors 4–6 times smaller than those of the baseline.

The third column of Fig. 4 shows results for partially cloudy scenes, defined here as total cloud cover between 10 %–90 %. In these cases, both schemes exhibit smaller errors than in fully cloudy scenes, consistent with the weaker overall cloud radiative impact. For the baseline, the bias is substantially smaller than fully cloudy conditions but still show a pronounced peak near 1 km for LW and a double peak at 1 and 10 km for SW. In contrast, the ML-enhanced samples produces nearly unbiased heating rates, with an MAE of  $0.082 \text{ K d}^{-1}$  (SW) and  $0.068 \text{ K d}^{-1}$  (LW), representing errors that are 5–10 times smaller than those of the baseline.

To interpret the double-peak error observed in partially cloudy samples, we divided the samples into precipitating and non-precipitating clouds, as a rough proxy for deep and shallow convection. Non-precipitating (warm) clouds were identified using thresholds of maximum  $0.01 \text{ mm h}^{-1}$ , total cloud cover larger than 10 %, and vertically integrated ice water path of less than  $1 \text{ g m}^{-2}$ . For reference, drizzle has a precipitation rate of  $0.2 \text{ mm d}^{-1}$  (Wood, 2012). Precipitating clouds were identified by total cloud cover > 10 % and precipitation more than  $3 \text{ mm h}^{-1}$ . For reference, Zhao et al. (2024) reports mean precipitation of  $3.5 \text{ mm h}^{-1}$  for deep convective cores over the tropical ocean. For non-precipitating clouds, both the coarse-scale radiation scheme and the ML-enhanced scheme show an MAE peak near 1 km for SW and LW. However, the ML-enhanced scheme achieves substantially smaller errors of  $0.080 \text{ K d}^{-1}$  for SW and  $0.069 \text{ K d}^{-1}$  for LW, which are 6–10 times lower than those of the baseline. For the precipitating clouds, the baseline exhibits an MAE peak at 10–12 km, while the ML-enhanced scheme shows enhanced error in the upper troposphere but without distinct peak. Instead, the ML-enhanced scheme shows a broader peak in MAE between 12–14 km. On average, however, the ML-enhanced error remain about 4–5 times smaller than those of the baseline.

We further evaluated the performance of the ML-enhanced scheme across five selected regions characterized by different predominant cloud regimes, following the classification of Bock and Lauer (2024). The results are shown in Fig. 5 and summarized in Table A2. In the arctic region ( $70^{\circ}$ – $90^{\circ}$  N), errors remain confined below 10 km consistent with the lower tropopause height in this region. For the baseline, the SW MAE is  $0.215 \text{ K d}^{-1}$ , even smaller than in the clear-sky conditions, although the spread in MAE is slightly larger. For LW, the MAE is  $0.614 \text{ K d}^{-1}$ , exceeding the clear-sky values. In contrast, the ML-enhanced scheme achieves errors that are 4–8 times smaller than those of the baseline.

In the Southern Ocean (30–65° S), the baseline exhibits large errors of  $0.417 \text{ K d}^{-1}$  for SW and  $0.760 \text{ K d}^{-1}$  for LW, with a characteristic double peak in the MAE at 1–2 and 10 km. The ML-enhanced scheme also reproduces this double peak structure in the MAE spread but reduces the MAE by a factor of 4–7 relative to the baseline. Over the tropical ocean (30° N–30° S), the baseline shows large errors in the upper troposphere, likely associated with deep convection. In this region, the ML-enhanced scheme again reduces the MAE by a factor of 5–9. A subregion within the tropical region, the Pacific ITCZ region (0–12° N, 135° E–85° W), shows similar behavior but with even larger errors at higher altitudes.

The stratocumulus region is represented by three subregions: the Southeast Pacific (10–30° N, 75–97° W), Southeast Atlantic (10–30° S, 10° W–10° E), and Northeast Pacific (15–35° N, 120–140° W). In these regions, both models show two peaks in the MAE: one in the lower troposphere at 1–2 km and another in the upper troposphere at 10–12 km. Notably, the upper tropospheric peak is larger in both SW and LW for both models. Nevertheless, the ML-enhanced scheme achieves errors 5–9 times smaller than those of the baseline.

## 5 Conclusions

ESMs struggle to represent subgrid-scale cloudiness and commonly rely on statistical schemes, such as McICA, to account for subgrid-scale cloud radiative effects (Pincus et al., 2003). Although random, unbiased errors can be mitigated by large-scale atmospheric mixing (Pincus et al., 2003), the column-by-column error can be large. ML algorithms trained on high-resolution, global storm-resolving simulations now provide an opportunity to represent fractional cloudiness in radiative transfer more accurately. To bridge scales, the high-resolution model output is coarse-grained to the target resolution, such that the coarse-grained variables implicitly retain subgrid-scale effects. Then, the coarse-grained variables implicitly include subgrid-scale effects.

We developed a hybrid physics-ML radiation parameterization, where the physics-based component computes clear-sky fluxes, while the ML component predicts cloud impact, implicitly accounting for subgrid-scale variability. This ML-enhanced framework offers a more robust and generalizable radiation scheme: the physics-based parameterization retains its responsiveness to changes in GHGs and aerosols, thereby mitigating potential out-of-distribution issues in climate projections. The ML component is implemented as a BiLSTM neural network, which has previously demonstrated strong performance in radiation applications (Ukkonen, 2022; Yao et al., 2023; Ukkonen and Chantry, 2025; Bertoli et al., 2025; Hafner et al., 2025b). For training, we use data from high-resolution QUBICC simulations with a horizontal resolution of  $\approx 5 \text{ km}$  and 191 vertical layers expanding up to 83 km.

These fields are coarse-grained to  $\approx 80 \text{ km}$  and 47 vertical layers, matching the target resolution for a coarse-scale ESM. For comparison and to assess systematic differences between high-resolution and coarse-scale models, we additionally perform a coarse-scale ICON-A simulation. The distributions of the relevant input variables are found to be comparable between the coarse-scale and coarse-grained simulations.

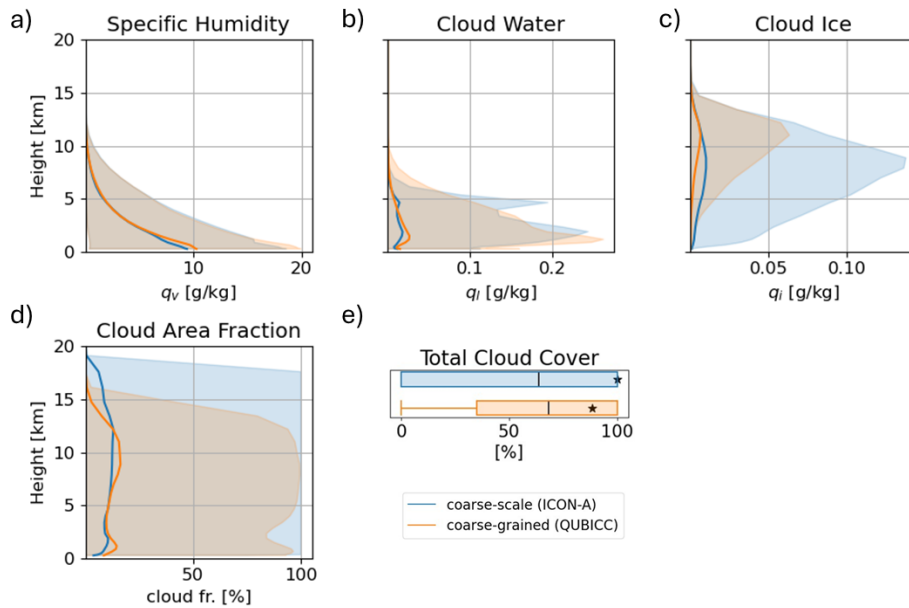
We find that a coarse-scale radiation scheme (baseline) performs well for clear-sky samples, but exhibits large errors in cloudy conditions, reflecting its inability to represent subgrid-scale distributions. In contrast, the ML-enhanced radiation scheme consistently outperforms the baseline, reducing errors from unresolved clouds in the radiative transfer calculations by a factor of 4–10. Although the ML-enhanced radiation scheme does not explicitly resolve subgrid-scale distributions, it learns how specific combinations of grid-scale mean states map to heating rates that implicitly include subgrid effects. The baseline showed substantial biases of  $1\text{--}3 \text{ K d}^{-1}$  in the upper troposphere at 10–15 km for precipitating clouds, highlighting the strong influence of subgrid-scale cloud ice on heating rates. In general, both coarse-scale radiation scheme and the ML-enhanced scheme produce larger errors in cloudier conditions, but the ML-enhanced scheme consistently yields smaller errors. These results emphasize the need for a more explicit treatment of subgrid-scale clouds, particularly in the upper troposphere.

Therefore, we conclude that high-resolution model data combined with ML can improve the representation of cloud–radiation interactions in coarse-scale radiation parameterizations. Nevertheless, the presented approach has caveats. High-resolution simulations at 5 km horizontal resolution cannot resolve shallow convection directly, leaving associated cloud radiative effects on heating rates – particularly within the planetary boundary layer – unresolved (Stevens et al., 2019). Using finer horizontal resolutions could help reduce these uncertainties. In addition, aerosols and heterogeneous GHG concentrations are typically absent from current high-resolution models. If future simulations include substantial variability in GHG and aerosol concentrations, these could be incorporated as additional NN inputs to capture secondary effects of reflected radiation.

One of the next steps is the online implementation of the ML-enhanced radiation scheme in a coarse-resolution model such as ICON-A. Although the online stability remains to be tested, the comparison with the coarse-scale model and results from previous stable hybrid simulations (Hafner et al., 2025a) are promising, suggesting potential improvement for climate projections. An additional advantage of the presented scheme is that clear-sky fluxes can be computed less frequently, while the cloud radiative impact can be updated every time step. This provides a pathway to both reducing computational costs and improving the representation of cloud–radiation interactions.

### Appendix A: Default microphysics scheme

One major difference between the ICON-A and QUBICC versions is the microphysics scheme. ICON-A uses a modified version of the Lohmann and Roeckner (1996) scheme and the QUBICC simulation uses the graupel scheme described in Doms et al. (2011). While both schemes are single-moment schemes, the latter treats precipitating tracers like snow, rain and graupel as prognostic variables while the former only diagnoses snow and rain. Moreover, it is known that cloud ice is too large in the upper troposphere in ICON-A (Doktorowski et al., 2024), which we also see when comparing cloud ice in ICON-A and QUBICC (Fig. A1).



**Figure A1.** As Fig. 2 but for the default microphysics scheme based on Lohmann and Roeckner (1996).

**Table A1.** Bulk statistics for heating rate results of the coarse-scale ML-based radiation emulator on coarse-grained QUBICC data. MAE is mean absolute error and  $R^2$  is coefficient of determination. RMSE is root mean squared error. The percentage values in parentheses denote the relative values of MAE, bias and RMSE.

	MAE [ $\text{K d}^{-1}$ ]	Bias [ $\text{K d}^{-1}$ ]	$R^2$	RMSE [ $\text{K d}^{-1}$ ]
pyRTE				
SW clear	0.367 (8.47 %)	0.234 (4.91 %)	0.91	0.443 (10.56 %)
SW cloudy	0.445 (16.73 %)	0.204 (4.22 %)	0.83	0.789 (41.54 %)
SW partial	0.470 (12.24 %)	0.273 (4.65 %)	0.82	0.683 (23.55 %)
SW non-precip. clouds	0.493 (12.11 %)	0.292 (4.79 %)	0.87	0.711 (19.98 %)
SW precip. clouds	0.778 (32.62 %)	0.244 (17.28 %)	0.59	1.250 (58.35 %)
LW clear	0.564 (23.56 %)	-0.349 (-9.85 %)	0.83	0.677 (30.09 %)
LW cloudy	0.862 (41.08 %)	-0.296 (-8.35 %)	0.67	1.478 (81.55 %)
LW partial	0.694 (25.83 %)	-0.328 (-7.74 %)	0.56	1.112 (47.24 %)
LW non-precip. clouds	0.725 (48.35 %)	-0.294 (-27.53 %)	0.70	1.230 (73.66 %)
LW precip. clouds	1.109 (77.76 %)	-0.337 (-23.85 %)	0.34	1.578 (130.42 %)
ML-enhanced				
SW clear	0.049 (0.49 %)	-0.000 (-0.00 %)	0.99	0.074 (0.72 %)
SW cloudy	0.106 (4.46 %)	0.012 (0.98 %)	0.98	0.214 (11.30 %)
SW partial	0.082 (2.00 %)	0.005 (0.23 %)	0.99	0.150 (5.00 %)
SW non-precip. clouds	0.080 (1.57 %)	0.006 (0.24 %)	0.99	0.144 (3.50 %)
SW precip. clouds	0.188 (9.16 %)	0.024 (2.87 %)	0.96	0.341 (17.71 %)
LW clear	0.028 (2.58 %)	0.003 (0.52 %)	1.00	0.053 (5.50 %)
LW cloudy	0.127 (7.42 %)	-0.001 (0.00 %)	0.98	0.275 (17.70 %)
LW partial	0.068 (3.33 %)	-0.001 (0.02 %)	0.99	0.158 (8.70 %)
LW non-precip. clouds	0.069 (4.43 %)	-0.002 (0.43 %)	0.99	0.160 (9.41 %)
LW precip. clouds	0.197 (19.37 %)	-0.004 (-0.99 %)	0.96	0.319 (36.18 %)

**Table A2.** Bulk statistics for results of cloud radiative effect on heating rates. MAE is the mean absolute error and  $R^2$  is the coefficient of determination. RMSE is the root mean squared error. The percentage values in parentheses denote the relative values of MAE, bias and RMSE.

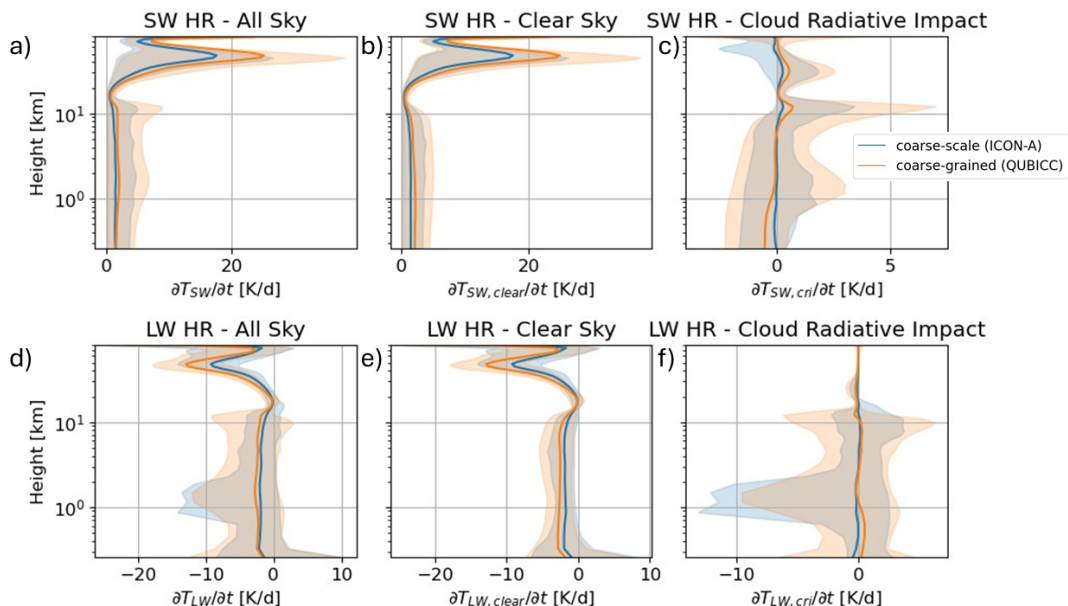
	MAE [ $\text{K d}^{-1}$ ]	Bias [ $\text{K d}^{-1}$ ]	$R^2$	RMSE [ $\text{K d}^{-1}$ ]
pyRTE				
SW Arctic	0.215 (9.55 %)	0.114 (3.73 %)	0.92	0.326 (17.99 %)
SW Southern Ocean	0.417 (13.34 %)	0.218 (3.32 %)	0.87	0.665 (27.68 %)
SW Tropics	0.535 (13.86 %)	0.290 (4.39 %)	0.80	0.817 (29.42 %)
SW Pacific ITCZ	0.638 (18.33 %)	0.271 (6.27 %)	0.76	0.964 (33.67 %)
SW stratocumulus	0.529 (13.66 %)	0.289 (4.38 %)	0.80	0.825 (28.97 %)
LW Arctic	0.634 (30.75 %)	-0.254 (-14.89 %)	0.77	1.046 (55.94 %)
LW Southern Ocean	0.787 (30.60 %)	-0.318 (-8.64 %)	0.69	1.307 (56.70 %)
LW Tropics	0.755 (30.96 %)	-0.326 (-4.21 %)	0.59	1.192 (65.09 %)
LW Pacific ITCZ	0.944 (50.80 %)	-0.311 (-2.19 %)	0.49	1.326 (99.80 %)
LW stratocumulus	0.742 (31.38 %)	-0.327 (-6.13 %)	0.58	1.234 (62.76 %)
ML-enhanced				
SW Arctic	0.056 (2.17 %)	0.006 (0.58 %)	0.99	0.098 (4.84 %)
SW Southern Ocean	0.089 (2.98 %)	0.005 (0.34 %)	0.99	0.174 (7.30 %)
SW Tropics	0.097 (2.61 %)	0.009 (0.47 %)	0.98	0.187 (7.32 %)
SW Pacific ITCZ	0.142 (4.50 %)	0.019 (1.17 %)	0.98	0.247 (9.73 %)
SW stratocumulus	0.095 (2.48 %)	0.008 (0.41 %)	0.99	0.179 (6.38 %)
LW Arctic	0.071 (4.96 %)	-0.000 (0.01 %)	0.99	0.176 (11.91 %)
LW Southern Ocean	0.103 (4.86 %)	-0.002 (-0.07 %)	0.99	0.231 (11.55 %)
LW Tropics	0.079 (4.67 %)	-0.002 (-0.02 %)	0.99	0.172 (14.14 %)
LW Pacific ITCZ	0.132 (10.55 %)	-0.002 (-0.46 %)	0.98	0.219 (23.47 %)
LW stratocumulus	0.080 (4.75 %)	-0.001 (0.11 %)	0.99	0.175 (11.59 %)

## Appendix B: Calculation of heating rates

When comparing the unscaled heating rates between coarse-scale and coarse-grained simulations, we find huge biases of up to  $10 \text{ K d}^{-1}$  between the mean heating rates, especially in the stratosphere. We found that there is a difference in heating rate calculation between the code versions. Usually, the heating rate is calculated from flux divergence, pressure difference and constants:

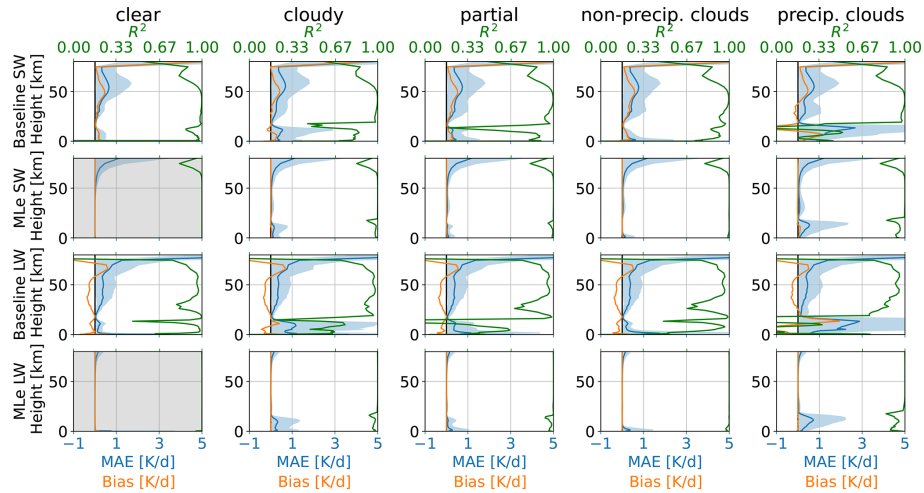
$$\frac{\partial T_k}{\partial t} = \frac{g}{c_p} \frac{F_{\text{Net},k+1/2} - F_{\text{Net},k-1/2}}{P_{k+1/2} - P_{k-1/2}}, \quad (\text{B1})$$

where  $g$  is gravitational acceleration,  $c_p$  specific heat at constant pressure,  $F_{\text{Net}}$  is the difference between downward and upward flux,  $P$  is pressure.  $k$  is defined at the center of a layer (also full levels) while  $k \pm \frac{1}{2}$  is defined at the layer boundaries (also half levels). This form of converting fluxes to heating rates is usually found in hydrostatic models but does not work in ICON because pressure is a diagnostic variable. Instead, the density is kept constant and specific heat at constant volume  $c_v$  needs to be used for the conversion (Zängl et al., 2014). This transforms Eq. (B1) to Eq. (2). In the ICON-A version, they use  $c_p$  which is valid for quasi-hydrostatic models because then the hydrostatic pressure holds  $dP/g = \rho dz$ . Additionally, the specific heat is scaled only by water vapor, while all tracers are included in the code version used for QUBICC. Therefore, we scaled the coarse-grained heating rates in Fig. 3 with the ratio  $c_v/c_p$ , which is on average 0.7. The unscaled heating rates are shown in Fig. B1.



**Figure B1.** As Fig. 3 but here the coarse-grained heating rates are not scaled.

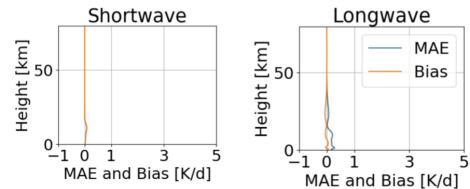
### Appendix C: Results for full vertical column



**Figure C1.** As Fig. 4 but for the full column. The large errors in the upper stratosphere for the baseline are related to rounding errors increasing with height.

### Appendix D: Linear decomposition under changing CO<sub>2</sub>

The linear decomposition assumption of clear-sky heating and cloud radiative impact may be questionable for different greenhouse gas concentrations. To provide some validity, we estimate the error induced by this assumption for  $4 \times \text{CO}_2$ . For a direct estimation, we select 4000 random samples from the test set and calculate the cloud radiative impact for the reference climate ( $\text{CO}_2$  concentration of 2004) offline using pyRTE. Next, we increase  $\text{CO}_2$  by a factor of 4 and repeat the calculation. The mean absolute difference between `cloud_effect_reference` and `cloud_effect_4xCO2` gives an estimate of the error that is induced by the linear decomposition assumption for  $4 \times \text{CO}_2$ . The vertical resolved error and bias are shown in Fig. D1. The  $x$ -range is the same as in Figs. 4, 5 and C1 for comparison with the errors induced by subgrid-scale clouds and the remaining errors of MLe-radiation. In general, the error is smaller in the stratosphere than in the troposphere. For SW, the error is around 2.5 % and for LW it is around 4 % in the troposphere. This around 10 times smaller than the error from subgrid-scale clouds and around the same magnitude as the errors from MLe-radiation.



**Figure D1.** Mean absolute error (MAE) and bias induced by the linear decomposition assumption for  $4 \times \text{CO}_2$ . For comparison, the  $x$ -range is the same as in Figs. 4, 5 and C1.

*Code and data availability.* The software code for the ICON model is available at <https://doi.org/10.35089/WDC/IconRelease01> (ICON Partnership et al., 2024) under the BSD-3C license. The exact version for high-resolution simulations is `icon-2024.10` and for coarse-resolution, the exact version (2.6.4) is archived at <https://doi.org/10.5281/zenodo.18853569> (Hafner, 2026). The ICON-model repository is archived under <https://doi.org/10.35089/WDC/IconRelease01> (ICON Partnership et al., 2024). The software code for pyRTE+RRTMGP is archived under <https://doi.org/10.5281/zenodo.16644555> (Pincus et al., 2025). The code for the network training and plots is archived under <https://doi.org/10.5281/zenodo.17280639> (Hafner, 2025). Data used for training and reproducing the plots is archived at <https://doi.org/10.5281/zenodo.18853569> (Hafner, 2026).

*Author contributions.* KH: Conceptualization, Investigation, Formal analysis, Software, Data Curation, Methodology, Writing – Original Draft. SS: Conceptualization, Methodology, Writing – Review and Editing. GB: Conceptualization, Methodology. AL: Supervision, Writing – Review and Editing. RP: Conceptualization, Software. JS: Data Curation, Writing – Review and Editing. VE: Funding acquisition, Supervision.

*Competing interests.* At least one of the (co-)authors is a member of the editorial board of *Geoscientific Model Development*. The peer-review process was guided by an independent editor, and the authors also have no other competing interests to declare.

*Disclaimer.* Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

*Acknowledgements.* Katharina Hafner and Veronika Eyring were supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the Gottfried Wilhelm Leibniz Prize awarded to Veronika Eyring (Reference No. EY 22/2-1). Veronika Eyring additionally acknowledge funding by the European Research Council (ERC) Synergy Grant “Understanding and Modeling the Earth System with Machine Learning” (US-MILE) under the Horizon 2020 Research and Innovation program (Grant Agreement No. 855187). This work used resources of the Deutsches Klimarechenzentrum (DKRZ) granted by its Scientific Steering Committee (WLA) under project ID bd1179. Robert Pincus, and Sara Shamekh were supported by the US National Science Foundation through the Learning the Earth with Artificial intelligence and Physics (LEAP) Science and Technology Center (STC) (Award #2019625). Sara Shamekh acknowledges support provided by Schmidt Sciences, LLC. The authors gratefully acknowledge the Earth System Modelling Project (ESM) for funding this work by providing computing time on the ESM partition of the supercomputer JUWELS at the Jülich Supercomputing Centre (JSC). Katharina Hafner acknowledges funding from the Swiss State Secretariat for Education, Research and Innovation (SERI) for the Horizon Europe project AI4PEX (Grant agreement ID: 101137682 and SERI no 23.00546). The authors would like to thank two anonymous reviewers that helped improving this manuscript.

*Financial support.* This research has been supported by the Deutsche Forschungsgemeinschaft (grant no. EY 22/2-1), the European Horizon 2020 (grant no. 855187), the Deutsches Klimarechenzentrum (grant no. bd1179), the National Science Foundation (grant no. 2019625), and the Jülich Supercomputing Centre, Forschungszentrum Jülich (grant no. icon-a-ml).

The article processing charges for this open-access publication were covered by the University of Bremen.

*Review statement.* This paper was edited by Po-Lun Ma and reviewed by two anonymous referees.

## References

- Barker, H. W., Stephens, G. L., and Fu, Q.: The sensitivity of domain-averaged solar fluxes to assumptions about cloud geometry, *Q. J. Roy. Meteor. Soc.*, 125, 2127–2152, <https://doi.org/10.1002/qj.49712555810>, 1999.
- Bertoli, G., Mohebi, S., Ozdemir, F., Jucker, J., Rüdüsühli, S., Perez-Cruz, F., Salzmann, M., and Schemm, S.: Revisiting Machine Learning Approaches for Short- and Longwave Radiation Inference in Weather and Climate Models, *J. Adv. Model. Earth Sy.*, 17, <https://doi.org/10.1029/2025ms004956>, 2025.
- Beucler, T., Gentine, P., Yuval, J., Gupta, A., Peng, L., Lin, J., Yu, S., Rasp, S., Ahmed, F., O’Gorman, P. A., Neelin, J. D., Lutsko, N. J., and Pritchard, M.: Climate-invariant machine learning, *Science Advances*, 10, eadj7250, <https://doi.org/10.1126/sciadv.adj7250>, 2024.
- Bock, L. and Lauer, A.: Cloud properties and their projected changes in CMIP models with low to high climate sensitivity, *Atmos. Chem. Phys.*, 24, 1587–1605, <https://doi.org/10.5194/acp-24-1587-2024>, 2024.
- Bony, S., Stevens, B., Frierson, D. M. W., Jakob, C., Kageyama, M., Pincus, R., Shepherd, T. G., Sherwood, S. C., Siebesma, A. P., Sobel, A. H., Watanabe, M., and Webb, M. J.: Clouds, circulation and climate sensitivity, *Nat. Geosci.*, 8, 261–268, <https://doi.org/10.1038/ngeo2398>, 2015.
- Bretherton, C. S., Henn, B., Kwa, A., Brenowitz, N. D., Watt-Meyer, O., McGibbon, J., Perkins, W. A., Clark, S. K., and Harris, L.: Correcting Coarse-Grid Weather and Climate Models by Machine Learning From Global Storm-Resolving Simulations, *J. Adv. Model. Earth Sy.*, 14, <https://doi.org/10.1029/2021ms002794>, 2022.
- Busecke, J. J. M., Balwada, D., Martin, P. E., Nicholas, T. E. G., Johnson, Z. C. P., Nalluri, P., Stern, C. I., and Abernathy, R. P.: The Impact of Sub-Grid Heterogeneity on Air-Sea Turbulent Heat Flux in Coupled Climate Models, *Geophys. Res. Lett.*, 52, <https://doi.org/10.1029/2025gl114951>, 2025.
- Chen, D., Rojas, M., Samset, B., Cobb, K., Diongue Niang, A., Edwards, P., Emori, S., Faria, S., Hawkins, E., Hope, P., Huybrechts, P., Meinshausen, M., Mustafa, S., Plattner, G.-K., and Tréguier, A.-M.: Framing, Context, and Methods, in: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., book section 1, pp. 147–286, Cambridge University Press, Cambridge, UK and New York, NY, USA, <https://doi.org/10.1017/9781009157896.003>, 2021.
- Chevallier, F., Chéruy, F., Scott, N. A., and Chédin, A.: A Neural Network Approach for a Fast and Accurate Computation of a Longwave Radiative Budget, *J. Appl. Meteorol. Clim.*, 37, 1385–1397, [https://doi.org/10.1175/1520-0450\(1998\)037<1385:ANNAFA>2.0.CO;2](https://doi.org/10.1175/1520-0450(1998)037<1385:ANNAFA>2.0.CO;2), 1998.

- Cotronei, A. and Slawig, T.: Single-precision arithmetic in ECHAM radiation reduces runtime and energy consumption, *Geosci. Model Dev.*, 13, 2783–2804, <https://doi.org/10.5194/gmd-13-2783-2020>, 2020.
- Doktorowski, S., Kretzschmar, J., Quaas, J., Salzmann, M., and Sourdeval, O.: Subgrid-scale variability of cloud ice in the ICON-AES 1.3.00, *Geosci. Model Dev.*, 17, 3099–3110, <https://doi.org/10.5194/gmd-17-3099-2024>, 2024.
- Doms, G., Förstner, G., Heise, E., Herzog, H.-J., Mironov, D., Raschendorfer, M., Reinhardt, T., Ritter, B., Schrodin, R., Schulz, J.-P., and Vogel, G.: A Description of the Nonhydrostatic Regional COSMO Model. Part II: Physical Parameterization, Consortium for Small-Scale Modelling, <http://www.cosmo-model.org/> (last access: 3 May 2026), 2011.
- Giorgetta, M. A., Brokopf, R., Crueger, T., Esch, M., Fiedler, S., Helmert, J., Hohenegger, C., Kornblueh, L., Köhler, M., Manzini, E., Mauritsen, T., Nam, C., Raddatz, T., Rast, S., Reinert, D., Sakradzija, M., Schmidt, H., Schneck, R., Schnur, R., Silvers, L., Wan, H., Zängl, G., and Stevens, B.: ICON-A, the Atmosphere Component of the ICON Earth System Model: I. Model Description, *J. Adv. Model. Earth Sy.*, 10, 1613–1637, <https://doi.org/10.1029/2017MS001242>, 2018.
- Giorgetta, M. A., Sawyer, W., Lapillonne, X., Adamidis, P., Alexeev, D., Clément, V., Dietlicher, R., Engels, J. F., Esch, M., Franke, H., Frauen, C., Hannah, W. M., Hillman, B. R., Kornblueh, L., Marti, P., Norman, M. R., Pincus, R., Rast, S., Reinert, D., Schnur, R., Schulzweida, U., and Stevens, B.: The ICON-A model for direct QBO simulations on GPUs (version icon-cscs:baf28a514), *Geosci. Model Dev.*, 15, 6985–7016, <https://doi.org/10.5194/gmd-15-6985-2022>, 2022.
- Grundner, A., Beucler, T., Gentine, P., Iglesias-Suarez, F., Giorgetta, M. A., and Eyring, V.: Deep Learning Based Cloud Cover Parameterization for ICON, *J. Adv. Model. Earth Sy.*, 14, e2021MS002959, <https://doi.org/10.1029/2021MS002959>, 2022.
- Hafner, K.: Representing Subgrid-Scale Cloud Effects in a Radiation Parameterization using Machine Learning: MLe-radiation v1.0, Zenodo [code], <https://doi.org/10.5281/ZENODO.17280639>, 2025.
- Hafner, K.: Code and Data for “Representing Subgrid-Scale Cloud Effects in a Radiation Parameterization using Machine Learning: MLe-radiation v1.0”, Zenodo [code and data set], <https://doi.org/10.5281/zenodo.18853569>, 2026.
- Hafner, K., Iglesias-Suarez, F., Shamekh, S., Gentine, P., Pincus, R., Giorgetta, M., and Eyring, V.: Stable Machine Learning based Radiation Emulation for ICON, *J. Adv. Model. Earth Sy.*, <https://doi.org/10.22541/essoar.174708082.27787580/v1>, in review, 2025a.
- Hafner, K., Iglesias-Suarez, F., Shamekh, S., Gentine, P., Pincus, R., Giorgetta, M., and Eyring, V.: Interpretable machine learning radiation parameterization for ICON, *J. Geophys. Res.-Mach. Learn. Comput.*, 2, e2024JH000501, <https://doi.org/10.1029/2024JH000501>, 2025b.
- Henn, B., Jauregui, Y. R., Clark, S. K., Brenowitz, N. D., McGibbon, J., Watt-Meyer, O., Pauling, A. G., and Bretherton, C. S.: A Machine Learning Parameterization of Clouds in a Coarse-Resolution Climate Model for Unbiased Radiation, *J. Adv. Model. Earth Sy.*, 16, <https://doi.org/10.1029/2023ms003949>, 2024.
- Heuer, H., Schwabe, M., Gentine, P., Giorgetta, M. A., and Eyring, V.: Interpretable Multiscale Machine Learning-Based Parameterizations of Convection for ICON, *J. Adv. Model. Earth Sy.*, 16, <https://doi.org/10.1029/2024ms004398>, 2024.
- Hogan, R. J. and Matricardi, M.: A Tool for Generating Fast k-Distribution Gas-Optics Models for Weather and Climate Applications, *J. Adv. Model. Earth Sy.*, 14, e2022MS003033, <https://doi.org/10.1029/2022MS003033>, 2022.
- Hohenegger, C., Korn, P., Linardakis, L., Redler, R., Schnur, R., Adamidis, P., Bao, J., Bastin, S., Behraves, M., Bergemann, M., Biercamp, J., Bockelmann, H., Brokopf, R., Brüggemann, N., Casaroli, L., Chegini, F., Datsaris, G., Esch, M., George, G., Giorgetta, M., Gutjahr, O., Haak, H., Hanke, M., Ilyina, T., Jahns, T., Jungclaus, J., Kern, M., Klocke, D., Kluft, L., Kölling, T., Kornblueh, L., Kosukhin, S., Kroll, C., Lee, J., Mauritsen, T., Mehlmann, C., Mieslinger, T., Naumann, A. K., Paccini, L., Peinado, A., Praturi, D. S., Putrasahan, D., Rast, S., Riddick, T., Roeber, N., Schmidt, H., Schulzweida, U., Schütte, F., Segura, H., Shevchenko, R., Singh, V., Specht, M., Stephan, C. C., von Storch, J.-S., Vogel, R., Wengel, C., Winkler, M., Ziemann, F., Marotzke, J., and Stevens, B.: ICON-Sapphire: simulating the components of the Earth system and their interactions at kilometer and subkilometer scales, *Geosci. Model Dev.*, 16, 779–811, <https://doi.org/10.5194/gmd-16-779-2023>, 2023.
- ICON Partnership (DWD, MPI-M, DKRZ, KIT, and C2SM): ICON release 2024.01, ICON Partnership [code], <https://doi.org/10.35089/WDC/IconRelease01>, 2024.
- Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, 3rd International Conference on Learning Representations, ICLR2015, San Diego, CA, USA, 7–9 May 2015, Conference Track Proceedings, <https://doi.org/10.48550/arXiv.1412.6980>, 2017.
- Klöwer, M., Gelbrecht, M., Hotta, D., Willmert, J., Silvestri, S., Wagner, G. L., White, A., Hatfield, S., Kimpson, T., Constantinou, N. C., and Hill, C.: SpeedyWeather.jl: Reinventing atmospheric general circulation models towards interactivity and extensibility, *Journal of Open Source Software*, 9, 6323, <https://doi.org/10.21105/joss.06323>, 2024.
- Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., Klöwer, M., Lottes, J., Rasp, S., Düben, P., Hatfield, S., Battaglia, P., Sanchez-Gonzalez, A., Willson, M., Brenner, M. P., and Hoyer, S.: Neural general circulation models for weather and climate, *Nature*, 632, 1060–1066, <https://doi.org/10.1038/s41586-024-07744-y>, 2024.
- Krasnopolsky, V. M., Fox-Rabinovitz, M. S., and Chalikov, D. V.: New Approach to Calculation of Atmospheric Model Physics: Accurate and Fast Neural Network Emulation of Longwave Radiation in a Climate Model, *Mon. Weather Rev.*, 133, 1370–1383, <https://doi.org/10.1175/MWR2923.1>, 2005.
- Lagerquist, R., Turner, D. D., Ebert-Uphoff, I., and Stewart, J. Q.: Estimating Full Longwave and Shortwave Radiative Transfer with Neural Networks of Varying Complexity, *J. Atmos. Ocean. Tech.*, 40, 1407–1432, <https://doi.org/10.1175/JTECH-D-23-0012.1>, 2023.
- LeCun, Y. A., Bottou, L., Orr, G. B., and Müller, K.-R.: Efficient BackProp, *Springer Berlin Heidelberg*, p. 9–48, [https://doi.org/10.1007/978-3-642-35289-8\\_3](https://doi.org/10.1007/978-3-642-35289-8_3), 2012.
- Lohmann, U. and Roeckner, E.: Design and performance of a new cloud microphysics scheme developed for the ECHAM

- general circulation model, *Clim. Dynam.*, 12, 557–572, <https://doi.org/10.1007/bf00207939>, 1996.
- Meyer, D., Hogan, R. J., Dueben, P. D., and Mason, S. L.: Machine Learning Emulation of 3D Cloud Radiative Effects, *J. Adv. Model. Earth Sy.*, 14, e2021MS002550, <https://doi.org/10.1029/2021MS002550>, 2022.
- Pal, A., Mahajan, S., and Norman, M. R.: Using Deep Neural Networks as Cost-Effective Surrogate Models for Super-Parameterized E3SM Radiative Transfer, *Geophys. Res. Lett.*, 46, 6069–6079, <https://doi.org/10.1029/2018GL081646>, 2019.
- Pincus, R., Barker, H. W., and Morcrette, J.: A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields, *J. Geophys. Res.-Atmos.*, 108, <https://doi.org/10.1029/2002jd003322>, 2003.
- Pincus, R., Mlawer, E. J., and Delamere, J. S.: Balancing Accuracy, Efficiency, and Flexibility in Radiation Calculations for Dynamical Models, *J. Adv. Model. Earth Sy.*, 11, 3074–3089, <https://doi.org/10.1029/2019MS001621>, 2019.
- Pincus, R., makepath LLC, and Sehnem, J. M.: pyRTE-RRTMGP, Zenodo [code], <https://doi.org/10.5281/zenodo.16644555>, 2025.
- Räsänen, P., Barker, H. W., Khairoutdinov, M. F., Li, J., and Randall, D. A.: Stochastic generation of subgrid-scale cloudy columns for large-scale models, *Q. J. Roy. Meteor. Soc.*, 130, 2047–2067, <https://doi.org/10.1256/qj.03.99>, 2004.
- Rasp, S.: Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations: general algorithms and Lorenz 96 case study (v1.0), *Geosci. Model Dev.*, 13, 2185–2196, <https://doi.org/10.5194/gmd-13-2185-2020>, 2020.
- Roh, S. and Song, H.-J.: Evaluation of Neural Network Emulations for Radiation Parameterization in Cloud Resolving Model, *Geophys. Res. Lett.*, 47, e2020GL089444, <https://doi.org/10.1029/2020GL089444>, 2020.
- Shamekh, S., Lamb, K. D., Huang, Y., and Gentine, P.: Implicit learning of convective organization explains precipitation stochasticity, *P. Natl. Acad. Sci. USA*, 120, <https://doi.org/10.1073/pnas.2216158120>, 2023.
- Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., Düben, P., Judt, F., Khairoutdinov, M., Klocke, D., Kodama, C., Kornbluh, L., Lin, S.-J., Neumann, P., Putman, W. M., Röber, N., Shibuya, R., Vanniere, B., Vidale, P. L., Wedi, N., and Zhou, L.: DYAMOND: the DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains, *Progress in Earth and Planetary Science*, 6, <https://doi.org/10.1186/s40645-019-0304-z>, 2019.
- Ukkonen, P.: Exploring Pathways to More Accurate Machine Learning Emulation of Atmospheric Radiative Transfer, *J. Adv. Model. Earth Syst.*, 14, e2021MS002875, <https://doi.org/10.1029/2021MS002875>, 2022.
- Ukkonen, P. and Chantry, M.: Vertically Recurrent Neural Networks for Sub-Grid Parameterization, *J. Adv. Model. Earth Sy.*, 17, e2024MS004833, <https://doi.org/10.1029/2024MS004833>, 2025.
- Ukkonen, P. and Hogan, R. J.: Twelve Times Faster yet Accurate: A New State-Of-The-Art in Radiation Schemes via Performance and Spectral Optimization, *J. Adv. Model. Earth Sy.*, 16, <https://doi.org/10.1029/2023ms003932>, 2024.
- Watt-Meyer, O., Brenowitz, N. D., Clark, S. K., Henn, B., Kwa, A., McGibbon, J., Perkins, W. A., Harris, L., and Bretherton, C. S.: Neural Network Parameterization of Subgrid-Scale Physics From a Realistic Geography Global Storm-Resolving Simulation, *J. Adv. Model. Earth Sy.*, 16, <https://doi.org/10.1029/2023ms003668>, 2024.
- Wood, R.: Stratocumulus Clouds, *Mon. Weather Rev.*, 140, 2373–2423, <https://doi.org/10.1175/mwr-d-11-00121.1>, 2012.
- Yao, Y., Zhong, X., Zheng, Y., and Wang, Z.: A Physics-Incorporated Deep Learning Framework for Parameterization of Atmospheric Radiative Transfer, *J. Adv. Model. Earth Sy.*, 15, e2022MS003445, <https://doi.org/10.1029/2022MS003445>, 2023.
- Zängl, G., Reinert, D., Rípodas, P., and Baldauf, M.: The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core, *Q. J. Roy. Meteor. Soc.*, 141, 563–579, <https://doi.org/10.1002/qj.2378>, 2014.
- Zhao, Y., Li, J., Wen, D., Li, Y., Wang, Y., and Huang, J.: Distinct structure, radiative effects, and precipitation characteristics of deep convection systems in the Tibetan Plateau compared to the tropical Indian Ocean, *Atmos. Chem. Phys.*, 24, 9435–9457, <https://doi.org/10.5194/acp-24-9435-2024>, 2024.