CAMBRIDGE
UNIVERSITY PRESS

APPLICATION PAPER

A physics-informed machine learning parameterization for cloud microphysics in ICON

Ellen Sarauer[1] , Mierk Schwabe[1] , Philipp Weiss[2], Axel Lauer[1], Philip Stier[2] and Veronika Eyring[1,3]

[1]Deutsches Zentrum für Luft und Raumfahrt e.V., Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[2]University of Oxford, Department of Physics, Atmospheric, Oceanic and Planetary Physics, Oxford, UK
[3]University of Bremen, Institute of Environmental Physics, Bremen, Germany
Corresponding author: Ellen Sarauer; Email: ellen.sarauer@dlr.de

Abstract

We developed a cloud microphysics parameterization for the icosahedral nonhydrostatic modeling framework (ICON) model based on physics-informed machine learning (ML). By training our ML model on high-resolution simulation data, we enhance the representation of cloud microphysics in Earth system models (ESMs) compared to traditional parameterization schemes, in particular by considering the influence of high-resolution dynamics that are not resolved in coarse ESMs. We run a global, kilometer-scale ICON simulation with a one-moment cloud microphysics scheme, the complex graupel scheme, to generate 12 days of training data. Our ML approach combines a microphysics trigger classifier and a regression model. The microphysics trigger classifier identifies the grid cells where changes due to the cloud microphysical parameterization are expected. In those, the workflow continues by calling the regression model and additionally includes physical constraints for mass positivity and water mass conservation to ensure physical consistency. The microphysics trigger classifier achieves an F1 score of 0.93 on classifying unseen grid cells. The regression model reaches an R^2 score of 0.72 averaged over all seven microphysical tendencies on simulated days used for validation only. This results in a combined offline performance of 0.78. Using explainability techniques, we explored the correlations between input and output features, finding a strong alignment with the graupel scheme and, hence, physical understanding of cloud microphysical processes. This parameterization provides the foundation to advance the representation of cloud microphysical processes in climate models with ML, leading to more accurate climate projections and improved comprehension of the Earth's climate system.

Impact Statement

This new cloud microphysics parameterization has the potential to enhance climate model accuracy, contributing to a better understanding of climate-relevant processes, in particular taking into account the influence of highly resolved dynamics on cloud microphysical processes. More accurate and robust climate projections are essential for informed decision-making and risk assessment in addressing the global warming crisis.

1. Introduction

The parameterization of cloud microphysics is a central component in Earth system models (ESMs), working closely coupled with the convection scheme to model the behavior of clouds in coarse-scale

This research article was awarded Open Materials badge for transparent practices. See the Data Availability Statement for details.

ESMs. It calculates the formation, growth, and removal of liquid and solid water particles and captures phase changes. The accurate representation of these processes is crucial for reliable climate projections, as they significantly influence the lifetime and brightness of clouds and, in general, the transport of water and energy through the atmosphere. As uncertainties in microphysical processes, particularly those involving ice particles, can greatly affect climate models, addressing these challenges is essential for improving climate projections (Morrison et al., 2020).

The focus of this study is to enhance the representation of subgrid-scale cloud microphysics within the coarse-scale icosahedral nonhydrostatic modeling framework (ICON) model (Zängl et al., 2014; Giorgetta et al., 2018), which traditionally employs the single-moment microphysical scheme of Lohmann and Roeckner (1996). This scheme solves prognostic equations for the mass mixing ratios (mmrs) of water vapor, cloud water, cloud ice, and rain. To improve upon this baseline, we develop a machine learning (ML)-based parameterization trained on high-resolution ICON simulation data that use a more complex single-moment scheme, that is, the graupel scheme of Baldauf et al. (2011), Rutledge and Hobbs (1984). Another option would have been to use a two-moment cloud microphysics scheme (Seifert and Beheng, 2006). These schemes provide a more detailed representation of cloud processes but are computationally significantly more expensive, especially for convection-permitting simulations at high resolution. Due to the computational constraints of our study, we chose a single-moment scheme. This scheme has been used in convection-permitting high-resolution simulations (simulation data for this study), as well as in coarse simulations together with an additional convection parameterization. In addition to the scheme of Lohmann and Roeckner (1996), it calculates the mmr of snow and graupel and provides precipitation rates for rain, snow, and graupel. All microphysical schemes considered here also solve the prognostic equations for temperature resulting from microphysical processes, such as latent heat release from phase changes.

ML made great strides in enhancing ESMs by replacing traditional parameterizations that are based on empirical and physical understanding and representing the statistical effect of a given process at the grid scale of the climate model (Rasp et al., 2018; Reichstein et al., 2019; Grundner et al., 2022; Fuchs et al., 2024). For this, ML models are trained on short high-resolution climate simulations or observations and then coupled to the coarse climate model, potentially eliminating long-standing biases in ESMs (Gentine et al., 2021; Eyring et al., 2024). Over the last years, learning the cloud microphysics parameterization with ML models has been an active field of research (Han et al., 2020; Gettelman et al., 2021; Perkins et al., 2023; Seifert and Siewert, 2024; Sharma and Greenberg, 2024). Existing ML approaches often struggle with the balance between computational efficiency and physical accuracy, particularly when upscaling from high-resolution data to coarser model grids.

So far, studies have emulated cloud microphysics parameterizations using the same model and resolution as the simulation data (Han et al., 2020; Gettelman et al., 2021; Perkins et al., 2023). These approaches aim to improve computational efficiency by accelerating ESMs while precisely replicating existing parameterizations, without expecting to enhance the physical representation of microphysics. Alternatively, some approaches learn from more detailed microphysics parameterizations using superdroplet box models (Seifert and Siewert, 2024; Sharma and Greenberg, 2024). These methods aim to enhance the representation of cloud microphysics and reduce biases in coarse-resolution ESMs through detailed high-resolution simulations. However, switching from superdroplet box models to global climate simulations can introduce issues. The differences in scale, boundary conditions, and dynamics between box models and ESMs can complicate the integration and accuracy of ML parameterizations. Furthermore, the computational requirements of ML-based superdroplet parameterizations differ significantly from classical parameterizations typically used in an ESM. This adds an extra layer of technical complexity when implementing such models in highly nonlinear systems like ICON.

In this study, we aimed to leverage the strengths of the previously mentioned approaches. We expand upon our previous work presented in Sarauer et al. (2024) and now provide a more comprehensive analysis. We use simulations from our target model, ICON, to generate training data for the ML parameterization. However, instead of merely emulating the model, we enhance our approach by learning from simulations with a better resolution at a convection-permitting horizontal scale of 5 km and use a more detailed single-moment scheme, that is, the graupel scheme. As a consequence, at the target

resolution of 80 km, the ML model is able to preserve the influence of crucial subgrid-scale dynamics and their nonlinear interactions with the larger-scale flow on cloud microphysics. Although this approach complicates the distinction between contributions from subgrid-scale dynamics and microphysical processes, it ensures a more accurate representation of their combined influence on cloud properties and their temporal evolution at coarser resolutions. This will be especially beneficial when parameterizing cloud convection in a combined parameterization with microphysics in future studies, where other dynamical parameters such as vertical updraft velocity or aerosol activation will be taken into account as additional input parameters. Even though the ML parameterization has not yet been coupled to ICON, this approach is believed to be able to enhance the representation of cloud microphysical processes in climate models, with the potential for more accurate climate projections by the resulting hybrid ESM (Eyring et al., 2024).

2. Methodology

2.1. Underlying dataset

We generate high-resolution data through a simulation setup utilizing the atmospheric component of the ESM ICON Sapphire (Hohenegger et al., 2023; Segura et al., 2025), in a similar setup to Weiss and Stier (2024). In contrast to the simulation of Hohenegger et al. (2023), the sea surface temperature and sea ice concentrations are prescribed. The horizontal resolution is about 5 km. The atmosphere is discretized with 90 vertical levels, and the land is discretized with 5 soil layers. The time step Δt is 40 s. We use ERA5 boundary conditions by Hersbach et al. (2020) and initialize the model with historical weather data, ozone concentrations, aerosol concentrations, ocean properties such as the sea surface temperature and sea ice concentration, and land properties. To keep the data volume manageable, we save the data every three simulated hours. Specifically, we store the time-averaged values of all input and output parameters used in the ML parameterization (see Appendix, Table A1) over 3-hour periods, rather than instantaneous tendencies. This allows the ML model to learn from smoothed data and reduces the likelihood of predicting extreme outliers that could destabilize online-coupled simulations (Yu et al., 2024). We run the simulation for 22 days, from 20 January to 9 February 2020, and discard the first 10 days as a spin-up period, which leaves us with 12 days of data. An illustration of the microphysical hydrometeors, used as inputs for our ML framework, derived from the simulation is given in Figure 1.

Subsequently, we apply horizontal coarse-graining mapping the data to a lower-resolution ICON grid (80 km) using the method from Grundner et al. (2022). The coarse-graining is performed on the entire dataset including both inputs (temperature, pressure, mmrs) and outputs (tendency temperature, tendencies mmrs). Coarse-graining is done by estimating grid-scale mean values through a weighted sum of high-resolution grid cells, where weights represent the fraction of the grid cell volume that each high-resolution cell fills. This allows us to approximate the large-scale state variables that are inputs for our cloud microphysics parameterization and ensures that the parameterization performs consistently across various spatial resolutions. After the coarse-graining, we select a random subset within the first 9 days after spin-up of our simulation for training and validation. We test our method on the 3 remaining days of the simulation. We scale the data with respect to the standard deviation and the mean of the training set.

2.2. Overview of ML model

Since the ultimate goal is to develop an ML-based cloud microphysics parameterization that runs online in an ICON simulation, we prioritize a multilayer perceptron (MLP)-based architecture developed in PyTorch (Paszke et al., 2019), which can be applied to each grid cell of the ICON model individually. Figure 2 shows an overview of the ML-based parameterization developed in this work. The considered input and output parameters are the same as in the traditional graupel scheme. The tendencies are solely based on microphysical processes, for example, tendency temperature considered here is only a result of latent heat release or cooling by phase changes. For the used set of input and output parameters and their description, the reader is referred to Appendix, Table A1. The ML-based cloud microphysics
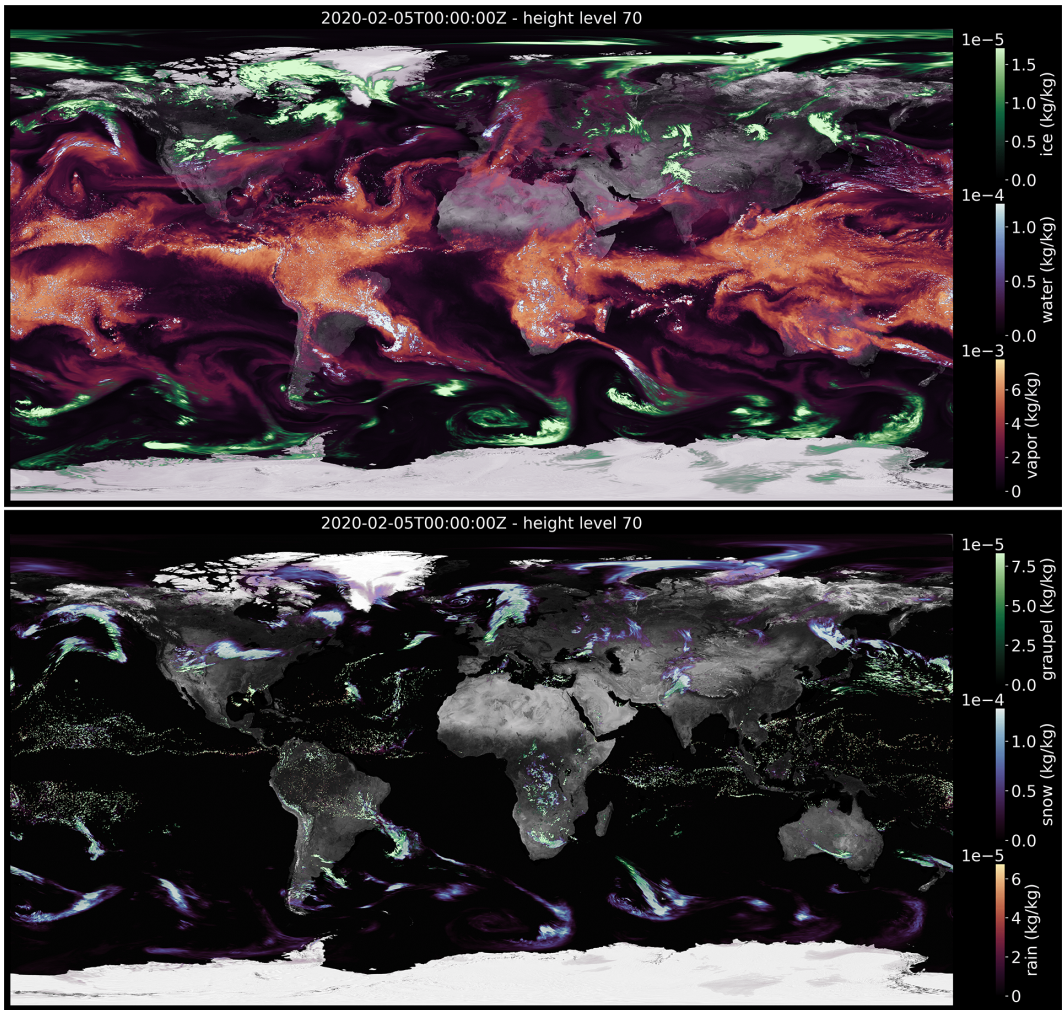
Figure 1. Map of mmrs. Obtained from the simulation on model level 70, corresponding to a height of about 3 km. The upper figure shows water vapor (red), cloud water (blue), and cloud ice (green). The lower figure shows rain (red), snow (blue), and graupel (green). Earth illustration from NASA Visible Earth (2024).

parameterization is split into a classifier and a regression task: if the classifier determines that in a specific grid cell, the microphysics scheme is triggered and the workflow continues by calling the regression model. Otherwise, we assume that all tendencies are zero for this grid cell and pass a zero vector. While end-to-end training of the classifier and regression tasks showed a similar performance as the modular approach, we opted for separate training to maintain modularity, which is crucial for the explainability of individual components. Nevertheless, for future work involving online coupling, an end-to-end trained model may be considered.

2.3. Microphysics trigger classifier

Similarly to Gettelman et al. (2021) and Perkins et al. (2023), we choose an approach to first classify the grid cells into those with and without active clouds, that is, according to whether we expect any change in our grid cell due to the cloud microphysical parameterization. Similar to Perkins et al. (2023), we need to apply proper preselection criteria to our dataset before the model training. Additionally, for the training
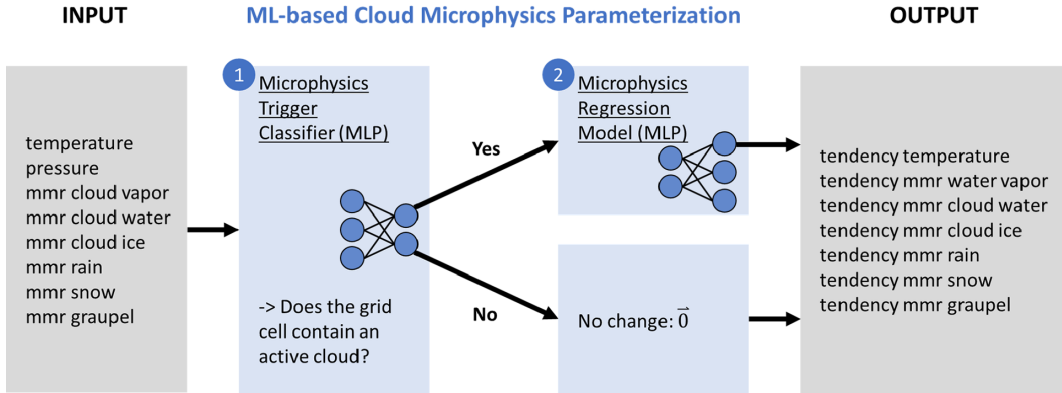
Figure 2. *Overview of the presented ML parameterization pipeline. Left, the input variables. Right, the output variables are listed, and for more details, see Table A1. In the center, step 1 illustrates the classifier MLP, and step 2 illustrates the regression MLP.*

and testing datasets, we only consider a grid cell as triggered when the sum of all microphysical hydrometeor tendencies Δm_q fulfills the criterion

$$\sum_q |\Delta m_q| \geq \epsilon, \tag{2.1}$$

with the computational threshold $\epsilon = 10^{-10}$. The tendency Δm_q represents the rate of change of the mmr per internal model time step Δt in the ICON simulation, which is 40 s in this study. By examining the underlying dataset of this study, we find that 81% of the grid cells are not changed by the microphysics scheme, which makes it especially important to sort out the inactive cells before performing the regression task. This happens in step 1 of our ML-based cloud microphysics parameterization. We train a simple MLP with 8 input nodes for the microphysical input parameters, as shown in Figure 2 with 2 hidden layers with 256 and 512 nodes and 1 output node for the classification task. We use dropout layers in between to avoid overfitting to the training dataset. The network uses rectified linear unit (ReLU) (Agarap, 2019) activation functions after the first and second layers to introduce nonlinearity and a sigmoid activation function at the end to produce a probability between 0 and 1. If the output value is larger than 0.5, we consider the microphysics scheme to be triggered and continue by calling the regression model. Otherwise, we assume that all tendencies are zero for this grid cell and pass a zero vector. We train the microphysics trigger classifier for 20 epochs on a random subset of 25 million samples selected from the 9 training days and evaluate on the 3 validation days.

2.4. Physics-constrained microphysics regression model

After training the trigger classifier, we predict the tendencies of triggered grid cells with the MLP-based regression model. It is structured as follows: 8 input nodes for the microphysical input parameters, 3 hidden layers with 512 nodes each using the ReLU activation (Agarap, 2019), and 7 output nodes using linear activation. Additionally, we include a residual connection between the input and output layers to preserve mass throughout the prediction for physical consistency. We use dropout layers and batch normalization to avoid overfitting to the training dataset. We train the microphysics regression model for 30 epochs on a random subset of 10 million samples selected from the 9 training days and evaluate on the 3 validation days.

During the preselection procedure of our simulated dataset, we find that the classical parameterization produces negative masses, or tendencies that lead to negative masses after the application of the classical parameterization in 4% of the samples in the preselected dataset. When these effects occur in the classical parameterization, the ICON model sets the negative masses to zero and recalculates the tendencies. To

incorporate this correction into our model, we apply the method of physics constraining for negative masses following the approach of Harder et al. (2022). We modify the weights of the output layer during the optimization of the MLP, constraining all tendencies Δq not fulfilling the condition

$$m_{q,t_1} = m_{q,t_0} + \Delta t \cdot \Delta m_q \geq 0. \tag{2.2}$$

where m_{q,t_1} is the mmr after the application of the parameterization, and m_{q,t_0} is the mmr before the application of the parameterization.

Additionally, we incorporate a constraint that ensures water mass conservation for the microphysical hydrometeors

$$\sum_q m_{q,t_1} = \sum_q \left(m_{q,t_0} + \Delta t \cdot \Delta m_q \right). \tag{2.3}$$

The total mmr of hydrometeors after applying the microphysics scheme, m_{q,t_1}, must equal the sum of mmrs of the initial hydrometeors m_{q,t_0} plus the sum of all tendencies Δm_q.

2.5. Explainability

As discussed by de Burgh-Day and Leeuwenburg (2023), incorporating domain knowledge into ML parameterizations increases the performance. Therefore, the strategy of this work is to stay as physically consistent as possible. Therefore, we introduce constraints to avoid unphysical negative mmrs. To further ensure consistency, we examine the model's behavior with explainability methods in each substep. For the purpose of explainability, we use the SHAP (SHapley Additive exPlanations package [Lundberg and Lee, 2017]). This package follows the game theory approach of the calculation of Shapley values. The Shapley values quantify the contribution of each input feature to the prediction of a model by comparing it to all possible combinations of features. The effect of this feature value on the output is determined by how much it increases or decreases the prediction relative to the model's baseline prediction, with the Shapley value reflecting this impact. We perform this analysis for the classifier and regression model individually. A high feature value in combination with a positive Shapley value means that with an increasing value of this input variable, we expect an increase in the respective model output. A low feature value and positive Shapley value means that by decreasing this feature value, we expect an increase in the output variable. The closer the Shapley value is to zero, the harder it is to explain the model's behavior. With these analyses, we can find the most relevant features for the two parts of our ML framework. Additionally, for each input feature, we can examine whether it has a negative or positive correlation with the model output. This helps us to see if the model is learning processes and relationships that we would actually expect from our physical understanding.

3. Results

Our study evaluates the performance of a physics-informed, ML-based cloud microphysics parameterization using coarse-grained data from a global, high-resolution ICON simulation for training, testing, and validation. We focus on identifying the MLP algorithms that best capture the data's characteristics while being as simple as possible for computational efficiency. This includes both the microphysics trigger classifier and the regression model.

3.1. Microphysics trigger classifier

For the first part of our ML framework, the microphysics trigger classifier, we develop an MLP-based classifier to predict whether a given grid cell state will result in changes to microphysical properties. The model achieves a good performance, with an F1 score of 0.93 and a receiver operating characteristic area under curve (ROC–AUC) (Fawcett, 2006) of 0.99. The F1 score measures a model's balance between precision and recall, while ROC–AUC assesses its overall ability to distinguish between classes.
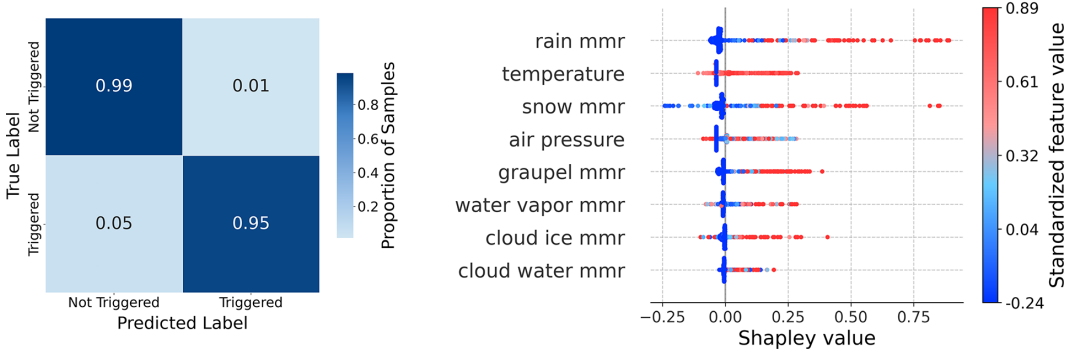
Microphysics trigger classifier

Figure 3. *Classification performance. The left image shows the confusion matrix, comparing predicted and true classes. The color intensity represents the ratio of samples in each class to the total number for that class. The right image shows the Shapley analysis. The input variables are listed in importance for the respective output variable from top to bottom. For each feature, red represents large values of a variable, and blue represents small ones. The x-axis shows the Shapley values.*

The performance is also illustrated in Figure 3. The confusion matrix shows the performance of the classification model by displaying the number of correct and incorrect predictions across different classes. The ratios on the diagonal tell the proportion of correctly classified samples out of the total actual samples of that class. The off-diagonal ratios represent the proportion of misclassified samples relative to the actual total in that class. We observe a good classification performance. Additionally, the similar ratios of false positives and false negatives indicate that the model does not exhibit a significant bias toward the dominant class, confirming the robustness of our classifier. Further analysis using explainability methods (Figure 3) shows that the mmr of rain is the most important feature for the microphysics trigger classifier. Across all mmrs of microphysical hydrometeors, lower values reduce the probability of microphysics triggering, while higher values increase it. This is consistent with our physical understanding, as more microphysical processes occur in regions with higher cloud water or precipitation content. Before implementing the classifier, we used threshold-based criteria to predict grid cell activation. Our Shapley plot supports these previous threshold methods, confirming that cells with sufficient cloud mmr (cloud water m_c and cloud ice m_i) or precipitation mmr (rain m_r, snow m_s, and graupel m_g) are more likely to activate the microphysics scheme, previously (Sarauer et al., 2024) defined by the following thresholds:

$$(m_c + m_i) \geq 10^{-5}\,\mathrm{kg \cdot kg^{-1}}, \quad (m_r + m_s + m_g) \geq 10^{-7}\,\mathrm{kg \cdot kg^{-1}}. \tag{3.1}$$

As expected, a broader range of features, such as temperature, air pressure, and water vapor concentration, are correlated with microphysics triggering, as microphysical processes occur across a wide range of values for these parameters and water vapor is present throughout the troposphere, where cloud microphysics takes place. This evaluation of the microphysics trigger classifier shows that the grid cells of interest are identified in a physically consistent manner. This provides a reliable basis for further analysis and thus can be used as a strong foundation for the subsequent regression tasks.

3.2. Microphysics regression model

In the second part of our ML framework, we address the regression task for the grid cells classified as triggered in the previous step. Despite the complexity of the cloud microphysics parameterization and low first-order correlations (cf. Appendix, Figure A1), we are able to reach an R^2 score of 0.72 averaged over

Table 1. *Fit performance measures for the regression output in the ML framework*

Variable	Description	Mean $\left[\frac{kg}{kg \cdot s}\right]$	$\sigma \left[\frac{kg}{kg \cdot s}\right]$	RMSE $\left[\frac{kg}{kg \cdot s}\right]$	R^2
tend_ta_mig	Tendency temperature	$3.87 \cdot 10^{-5}$	$2.31 \cdot 10^{-4}$	$9.85 \cdot 10^{-5}$	0.685
tend_qv_mig	Tendency mmr water vapor	$1.77 \cdot 10^{-8}$	$9.91 \cdot 10^{-8}$	$4.20 \cdot 10^{-8}$	0.687
tend_qc_mig	Tendency mmr cloud water	$8.51 \cdot 10^{-10}$	$8.92 \cdot 10^{-8}$	$4.01 \cdot 10^{-8}$	0.608
tend_qi_mig	Tendency mmr cloud ice	$2.71 \cdot 10^{-10}$	$2.92 \cdot 10^{-9}$	$1.39 \cdot 10^{-9}$	0.837
tend_qr_mig	Tendency mmr rain	$1.13 \cdot 10^{-9}$	$1.71 \cdot 10^{-8}$	$7.90 \cdot 10^{-9}$	0.739
tend_qs_mig	Tendency mmr snow	$5.46 \cdot 10^{-10}$	$8.91 \cdot 10^{-9}$	$6.02 \cdot 10^{-9}$	0.646
tend_qg_mig	Tendency mmr graupel	$5.02 \cdot 10^{-10}$	$1.86 \cdot 10^{-8}$	$8.44 \cdot 10^{-9}$	0.837

Note. The table shows for each output variable: arithmetic average (mean) and standard deviation (σ), root mean squared error (RMSE), and coefficient of determination (R^2).

all output features, achieving the best regression output for the tendencies of cloud ice and graupel mmr (both 0.84 R^2 score) and the worst regression output for the tendency of cloud water mmr (0.61 R^2 score). The specific goodness of fit for the individual output features is presented in Table 1.

The low score for cloud water in the model likely stems from its complex and sensitive nature, driven by many nonlinear, small-scale processes like droplet formation and growth, which are difficult to accurately parameterize. Cloud water varies greatly at smaller scales, making it harder to represent accurately. Additionally, the multitude of interactions of cloud water with other microphysical variables can complicate an accurate prediction. On the other hand, cloud ice, graupel, snow, and rain are easier to learn for the ML model, likely because their process representation is simpler in the classical, single moment graupel scheme. Moreover, variables such as temperature and water vapor are typically less sensitive, resulting in better model performance for these variables.

As illustrated in Figure 4, we see a good alignment between the prediction and test datasets (ground truth) for cloud ice. A similar illustration for the other variables can be found in Appendix, Figures A2 and A3. For all output features, the model effectively represents values near zero but has difficulty capturing extreme outliers. Despite experimenting with various scaling methods such as the Standard Scaler,
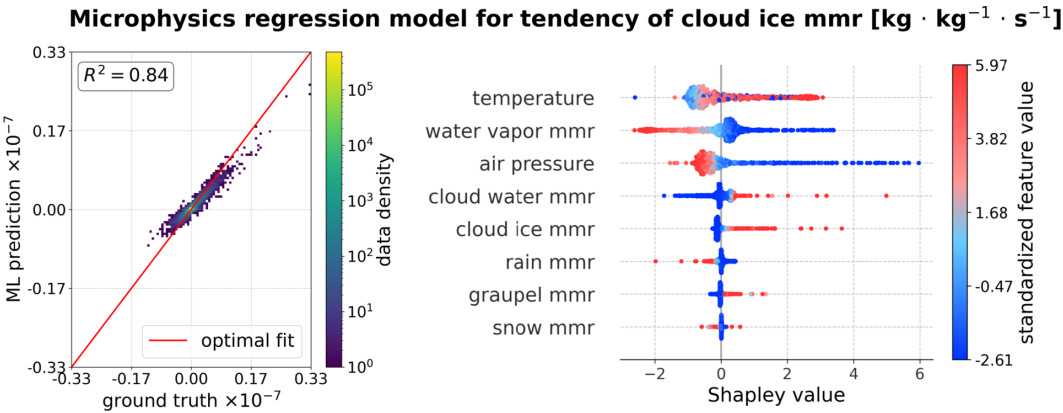

Figure 4. *Regression performance. The left image shows the ML predictions of the regression model versus ground truth of the tendency of cloud ice mmr. The colors illustrate the density of the data on a logarithmic scale. The right image shows the Shapley analysis for the tendency of cloud ice mmr. The input variables are listed in importance for the respective output variable from top to bottom. For each feature, red represents large values of a variable, and blue represents small ones. The x-axis shows the Shapley values.*

MinMax Scaler, Robust Scaler, and Log Scaler, accurately capturing these distributions remains a challenge for ML algorithms, as none of the approaches investigated fully captured the variability in the data while maintaining optimal training performance. In this work, we use the Standard Scaler, a commonly applied preprocessing technique for training such models (e.g., Heuer et al., 2024). To gain more insight into the relationships between input features and resulting model output, that is, microphysical tendencies, we apply explainability techniques.

Figure 4 shows illustrations of our explainability method for the microphysical tendency of cloud ice. The Shapley illustrations of the other tendencies can be found in Appendix, Figures A2 and A3. Generally, across all features, we observe a strong correlation between microphysical tendencies and both air pressure and temperature. This correlation is expected, as water vapor is of great importance to cloud formation and evolution, which itself is strongly influenced by temperature and pressure. The cloud ice tendency is positively correlated with a higher amount of cloud water and cloud ice content, as indicated by positive Shapley values for positive feature values. This could be a result of the frequent coexistence of cloud water and cloud ice in mixed clouds. Conversely, high feature values for water vapor go together with mostly negative Shapley values. Additionally, we would expect with our physical understanding that low temperature values should have positive Shapley values for ice, as low temperatures are required for the formation of cloud ice. However, the ML model shows positive Shapley values for cloud ice mmr for high temperature values. This could be due to higher temperatures resulting in a higher amount of water vapor that can, for example, condensate, which can lead to higher microphysical tendencies.

Constraining unphysical values during the training through a modified loss function, prohibiting negative masses, and ensuring water mass conservation as described above does not have an impact on the overall performance of the ML parameterization. Nevertheless, we expect the constrained ML parameterization to lead to an improved and stable simulation when coupled with a climate model in future work.

3.3. Combined model

To evaluate the overall performance of our ML-based cloud microphysics parameterization, we combine the results of the microphysics trigger classifier and the physics-informed regression model. In this study, the classifier achieved an F1 score of 0.93. This is particularly important because the dataset is imbalanced, with only 19% of the grid cells with triggered cloud microphysics processes. For the regression task, the model achieved an average R^2 score of 0.72. This R^2 score reflects the regression model's ability to predict the continuous outputs related to cloud microphysics processes within the triggered cells. To provide a unified measure of performance and to allow for a fair comparison with column-based methods like Perkins et al. (2023), we introduce a unified R^2 metric. This approach treats the classification of non-triggered cells as regression outputs mapped to zero. Specifically, for all grid cells, we compare the predicted values, which combine the classifier's predictions and the regression model's outputs, against the true values, where nontriggered cells have an expected output of 0. This unified approach ensures that the metric reflects performance across both tasks, including cases where the regressor is incorrectly activated. This calculation results in an overall score of 0.78, which reflects a strong model performance, balancing high classification accuracy with solid regression precision.

To conclude the analysis of our results, we consider the complex graupel mmr tendency predicted by the ML model in comparison with the coarse-grained high-resolution data and a coarse reference simulation (Figure 5). Similar illustrations for the microphysical tendencies can be found in Appendix, Figures A4 and A5. The white spaces in the figures indicate grid cells where the microphysics scheme is not activated, which illustrates the good representation of the triggered cells by the microphysics trigger classifier. Nevertheless, when looking at the predicted values with the regression, we see an overestimation of the tendency of graupel mmr, especially in the subtropics compared to the ground truth. Additionally, we observe biases in the ML model for some extreme values of the other tendencies, for example, underestimating the extremes in the tendency of cloud ice mmr and overestimating the extremes in the tendency of cloud water mmr. This could also introduce discrepancies when coupling the model

ML prediction, ground truth and baseline for tendency graupel [kg · kg⁻1 · s⁻1]
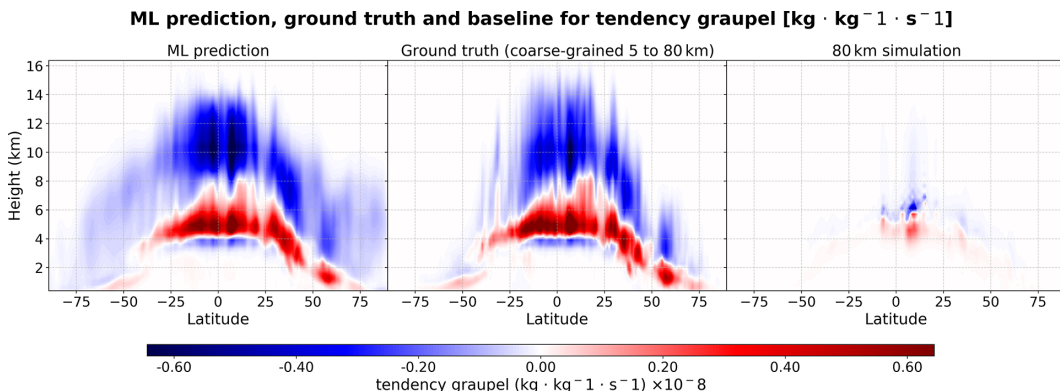

Figure 5. ML prediction (left) of the tendency of graupel mmr averaged over 3 validation days (24 timesteps in total) compared with the ground truth (center) and a coarse reference simulation (right) for a random day within the validation period. Colors represent the magnitude of the mmr tendency, averaged over longitude. The x-axis is latitude, the y-axis is height, and white areas indicate no change due to microphysics.

with a coarse global climate model. Appendix includes the distributions of these tendencies and a brief discussion. Despite this, the model demonstrates a reasonable distinction between areas of positive and negative tendencies. It is important to highlight that our model achieves these results using only cell-based information. This makes the model's performance particularly interesting, as it successfully captures physical processes and spatial distinctions. We compare our results with a reference simulation performed at a coarser horizontal resolution. For this, we use the same setup as before and coarse-grain all initial inputs for the simulation and run the simulation on the 80-km grid. This comparison shows that graupel processes are much less pronounced in comparison to the high-resolution simulation, even after the application of coarse-graining. However, since a convection parameterization is typically used at these coarse resolutions, we expect to see less graupel. These findings support our hypothesis that our ML-based cloud microphysics parameterization has the potential to improve the representation of microphysical processes by implicitly better resolving relevant dynamics and variability compared to an approach purely based on coarse-resolution output when integrated into the ICON model.

4. Conclusions

In this study, we conduct an analysis of cloud microphysical processes using a 22-day high-resolution ICON simulation. Based on a simulation with 5-km resolution with 90 vertical layers and coarse-graining of the data to 80-km resolution, we develop a novel ML-based framework to parameterize cloud microphysics. Our approach combines a physics-informed MLP algorithm, which includes a microphysics trigger classifier and a regression model, to address the complex problem of cloud microphysics. The use of a high-resolution model simulation and a cell-based approach provides valuable insights into the microphysical processes by learning from better resolved atmospheric dynamics that cannot be explicitly resolved in low-resolution models.

The microphysics trigger classifier, which achieved an F1 score of 0.93 and an ROC–AUC of 0.99, demonstrates good performance in identifying grid cells where microphysical processes are likely to occur. This strong classification performance is important to select 19% of the grid cells that are influenced by cloud microphysics processes. The classifier's ability to accurately identify these cells ensures that our regression model is applied only to the relevant cells. This will be especially important when coupling this model to an ESM online.

The regression model, trained on the identified triggered grid cells, achieved an average R^2 score of 0.72 across all output features. In particular, the model performs well in predicting the tendencies of variables such as cloud ice and graupel, with the highest performance observed for both (0.84 R^2). However, challenges remain with the mmr of cloud water, which could stem from its representation in the single moment graupel scheme and that cloud water is affected by a multitude of complex, small-scale processes and therefore challenging to predict.

Our analysis aims for physical consistency using explainability methods. The Shapley values indicate that features related to cloud water and cloud ice are the most important quantities in determining microphysics triggering. This aligns with our physical understanding, where the presence of higher cloud water or precipitation content increases the likelihood of phase changes due to microphysical processes. The combined performance of our ML models, with an overall performance score of 0.78, shows that the combination of classification and regression is effective. The model is able to balance classification accuracy with regression precision. The combined model's prediction of graupel mmr tendencies, while demonstrating an overestimation in certain regions, still captures important spatial distinctions and provides a promising approach to cloud microphysics parameterization.

While our results are promising, generating a more balanced training and validation dataset through additional simulations at various times of the year could further enhance model performance. Furthermore, vertical updrafts, which are critical for accurately capturing some of the cloud microphysical key processes, are only partially resolved even at this resolution. Updrafts influence microphysical rates such as condensation, evaporation, and precipitation formation and are a significant source of biases in coarse-resolution microphysical schemes where these dynamics are entirely parameterized (Stevens et al., 2020). Our approach bridges this gap by indirectly learning the effects of unresolved dynamics on microphysics from the high-resolution simulations, enabling the ML parameterization to some of these even when applied to coarse-scale models. Future improvements could focus on utilizing even higher-resolution data, such as those from Large Eddy Simulations (LES), which resolve vertical motion and mixing even more accurately. Such advancements could further reduce biases and enhance the physical accuracy of ML-based parameterizations for cloud microphysics. Finally, while the ML-based microphysics parameterization has not yet been coupled to ICON, implementing it in a global climate model will be crucial for future work to fully assess potential benefits and impacts on climate projections. This step poses significant technical challenges, including instabilities in simulation runs, the need for model retuning, and the complexity of Fortran-Python bridges. Addressing these points is an ongoing effort, and the technical details will be documented in future work. Overall, this study presents a novel approach in parameterizing cloud microphysics, contributing to a better understanding of climate-relevant processes and with the potential to improve the accuracy and robustness of climate projections, once coupled with other ML-based parameterizations to advanced hybrid ESMs (Eyring et al., 2024).

References

Agarap AF (2019) Deep learning using rectified linear units (ReLU). https://arxiv.org/abs/1803.08375

Baldauf M, **Seifert A**, **Förstner J**, **Majewski D**, **Raschendorfer M and Reinhardt T** (2011) Operational convective-scale numerical weather prediction with the cosmo model: description and sensitivities. *Monthly Weather Review 139*(12), 3887–3905. https://doi.org/10.1175/MWR-D-10-05013.1.

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

Eyring V, **Collins WD**, **Gentine P**, **Barnes EA**, **Barreiro M**, **Beucler T**, **Bocquet M**, **Bretherton CS**, **Christensen HM**, **Dagon K**, **Gagne DJ**, **Hall D**, **Hammerling D**, **Hoyer S**, **Iglesias-Suarez F**, **Lopez-Gomez I**, **McGraw MC**, **Meehl GA**, **Molina MJ and Zanna L** (2024) Pushing the frontiers in climate modelling and analysis with machine learning. *Nature Climate Change 14*, 1–13. https://doi.org/10.1038/s41558-024-02095-y.

Fawcett T (2006) Introduction to ROC analysis. *Pattern Recognition Letters 27*, 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.

Fuchs D, **Sherwood SC**, **Prasad A**, **Trapeznikov K and Gimlett J** (2024) Torchclim v1.0: a deep-learning plugin for climate model physics. *Geoscientific Model Development 17*(14), 5459–5475. https://doi.org/10.5194/gmd-17-5459-2024.

Gentine P, **Eyring V and Beucler T** (2021) Deep learning for the parametrization of subgrid processes in climate models. In *Deep Learning for the Earth Sciences*. Chichester, UK: John Wiley & Sons, Ltd, pp. 307–314. https://doi.org/10.1002/9781119646181.ch21.

Gettelman A, **Gagne DJ**, **Chen CC**, **Christensen MW**, **Lebo ZJ**, **Morrison H and Gantos G** (2021) Machine learning the warm rain process. *Journal of Advances in Modeling Earth Systems 13*(2), e2020MS002268. https://doi.org/10.1029/2020MS002268.

Giorgetta MA, **Sawyer W**, **Lapillonne X**, **Adamidis P**, **Alexeev D**, **Clément V**, **Dietlicher R**, **Engels JF**, **Esch M**, **Franke H**, **Frauen C**, **Hannah WM**, **Hillman BR**, **Kornblueh L**, **Marti P**, **Norman MR**, **Pincus R**, **Rast S**, **Reinert D**, **Schnur R**, **Schulzweida U and Stevens B** (2018) Icon-a, the atmosphere component of the icon earth system model: i model description. *Journal of Advances in Modeling Earth Systems 10*(7), 1613–1637. https://doi.org/10.1029/2017MS001242.

Grundner A, **Beucler T**, **Gentine P**, **Iglesias-Suarez F**, **Giorgetta MA and Eyring V** (2022) Deep learning based cloud cover parameterization for icon. *Journal of Advances in Modeling Earth Systems 14*(12), e2021MS002959. https://doi.org/10.1029/2021MS002959.

Han Y, **Zhang GJ**, **Huang X and Wang Y** (2020) A moist physics parameterization based on deep learning. *Journal of Advances in Modeling Earth Systems 12*(9), e2020MS002076. https://doi.org/10.1029/2020MS002076.

Harder P, **Watson-Parris D**, **Stier P**, **Strassel D**, **Gauger NR and Keuper J** (2022) Physics-informed learning of aerosol microphysics. *Environmental Data Science 1*, e20. https://doi.org/10.1017/eds.2022.22.

Hersbach H, **Bell B**, **Berrisford P**, **Hirahara S**, **Horányi A**, **Muñoz-Sabater J**, **Nicolas J**, **Peubey C**, **Radu R**, **Schepers D**, **Simmons A**, **Soci C**, **Abdalla S**, **Abellan X**, **Balsamo G**, **Bechtold P**, **Biavati G**, **Bidlot J**, **Bonavita M**, **De Chiara G**, **Dahlgren P**, **Dee D**, **Diamantakis M**, **Dragani R**, **Flemming J**, **Forbes R**, **Fuentes M**, **Geer A**, **Haimberger L**, **Healy S**, **Hogan RJ**, **Hólm E**, **Janisková M**, **Keeley S**, **Laloyaux P**, **Lopez P**, **Lupu C**, **Radnoti G**, **de Rosnay P**, **Rozum I**, **Vamborg F**, **Villaume S and Thépaut J** (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society 146*(730), 1999–2049. https://doi.org/10.1002/qj.3803.

Heuer H, **Schwabe M**, **Gentine P**, **Giorgetta MA and Eyring V** (2024) Interpretable multiscale machine learning-based parameterizations of convection for ICON. *Journal of Advances in Modeling Earth Systems 16*(8). https://doi.org/10.1029/2024ms004398.

Hohenegger C, **Korn P**, **Linardakis L**, **Redler R**, **Schnur R**, **Adamidis P**, **Bao J**, **Bastin S**, **Behravesh M**, **Bergemann M**, **Biercamp J**, **Bockelmann H**, **Brokopf R**, **Brüggemann N**, **Casaroli L**, **Chegini F**, **Datseris 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 AK**, **Paccini L**, **Peinado A**, **Praturi DS**, **Putrasahan D**, **Rast S**, **Riddick T**, **Roeber N**, **Schmidt H**, **Schulzweida U**, **Schütte F**, **Segura H**, **Shevchenko R**, **Singh V**, **Specht M**, **Stephan CC**, **von Storch J-S**, **Vogel R**, **Wengel C**, **Winkler M**, **Ziemen F**, **Marotzke J and Stevens B** (2023) Icon-Sapphire: simulating the components of the earth system and their interactions at kilometer and subkilometer scales. *Geoscientific Model Development 16*(2), 779–811. https://doi.org/10.5194/gmd-16-779-2023.

Jülich Supercomputing Centre (2021) JUWELS cluster and booster: exascale pathfinder with modular supercomputing architecture at Juelich Supercomputing Centre. *Journal of Large-Scale Research Facilities 7*(A138). https://doi.org/10.17815/jlsrf-7-183.

Lohmann U and Roeckner E (1996) Design and performance of a new cloud microphysics scheme developed for the ECHAM general circulation model. *Climate Dynamics 12*, 557–572.

Lundberg SM and Lee S (2017) A unified approach to interpreting model predictions. In Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S and Garnett R (eds.), *Advances in Neural Information Processing Systems 30*. Red Hook, NY: Curran Associates, Inc, pp. 4765–4774.

Morrison H, **van Lier-Walqui M**, **Fridlind AM**, **Grabowski WW**, **Harrington JY**, **Hoose C**, **Korolev A**, **Kumjian MR**, **Milbrandt JA**, **Pawlowska H**, **Posselt DJ**, **Prat OP**, **Reimel KJ**, **Shima S**, **van Diedenhoven B and Xue L** (2020) Confronting the challenge of modeling cloud and precipitation microphysics. *Journal of Advances in Modeling Earth Systems 12*(8), e2019MS001689. https://doi.org/10.1029/2019MS001689.

NASA Visible Earth (2024) NASA Visible Earth. (accessed 22 January 2024). https://visibleearth.nasa.gov/

Paszke A, **Gross S**, **Massa F**, **Lerer A**, **Bradbury J**, **Chanan G**, **Killeen T**, **Lin Z**, **Gimelshein N**, **Antiga L**, **Desmaison A**, **Köpf A**, **Yang E**, **DeVito Z**, **Raison M**, **Tejani A**, **Chilamkurthy S**, **Steiner B**, **Fang L**, **Bai J and Chintala S** (2019) Pytorch: an Imperative Style, High-Performance Deep Learning Library, 8024–8035. Available at http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

Perkins WA, **Brenowitz ND**, **Bretherton CS and Nugent JM** (2023) Emulation of cloud microphysics in a climate model. *Journal of Advances in Modeling Earth Systems 16*, e2023MS003851. https://doi.org/10.1029/2023MS003851

Rasp S, **Pritchard MS and Gentine P** (2018) Deep learning to represent sub-grid processes in climate models [arXiv:1806.04731 [physics, stat]]. *Proceedings of the National Academy of Sciences 115*(39), 9684–9689. https://doi.org/10.1073/pnas.1810286115.

Reichstein M, **Camps-Valls G**, **Stevens B**, **Jung M**, **Denzler J**, **Carvalhais N and Prabhat** (2019) Deep learning and process understanding for data-driven earth system science. *Nature 566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Rutledge SA and Hobbs PV (1984) The mesoscale and microscale structure and organization of clouds and precipitation in midlatitude cyclones. xii: a diagnostic modeling study of precipitation development in narrow cold-frontal rainbands. *Journal of Atmospheric Sciences 41*(20), 2949–2972. https://doi.org/10.1175/1520-0469(1984)041<2949:TMAMSA>2.0.CO;2.

Sarauer E, **Schwabe M**, **Lauer A**, **Stier P**, **Weiss P and Eyring V** (2024) Physics-informed machine learning-based cloud microphysics parameterization for earth system models. *International Conference on Learning Representations 2024*.

Segura H, **Pedruzo-Bagazgoitia X**, **Weiss P**, **Muller SK**, **Rackow T**, **Lee J**, **Dolores-Tesillos E**, **Benedict I**, **Aengenheyster M**, **Aguridan R**, **Arduini G**, **Baker AJ**, **Bao J**, **Bastin S**, **Baulenas E**, **Becker T**, **Beyer S**, **Bockelmann H**, **Bruggemann N**, **Brunner L**, **Cheedela SK**, **Das S**, **Denissen J**, **Dragaud I**, **Dziekan P**, **Ekblom M**, **Engels JF**, **Esch M**, **Forbes R**, **Frauen C**, **Freischem L**, **Garcia-Maroto D**, **Geier P**, **Gierz P**, **Gonzalez-Cervera A**, **Grayson K**, **Griffith M**, **Gutjahr O**, **Haak H**, **Hadade I**, **Haslehner K**, **ul Hasson S**, **Hegewald J**, **Kluft L**, **Koldunov A**, **Koldunov N**, **Kolling T**, **Koseki S**, **Kosukhin S**, **Kousal J**, **Kuma P**, **Kumar AU**, **Li R**, **Maury N**, **Meindl M**, **Milinski S**, **Mogensen K**, **Niraula B**, **Nowak J**, **Praturi DS**, **Proske U**, **Putrasahan D**, **Redler R**, **Santuy D**, **Sarmany D**, **Schnur R**, **Scholz P**, **Sidorenko D**, **Spat D**, **Sutzl B**, **Takasuka D**, **Tompkins A**, **Uribe A**, **Valentini M**, **Veerman M**, **Voigt A**, **Warnau S**, **Wachsmann F**, **Waclawczyk M**, **Wedi N**, **Wieners K-H**, **Wille J**, **Winkler M**, **Wu Y**, **Ziemen F**, **Zimmermann J**, **Bender FA-M**, **Bojovic D**, **Bony S**, **Bordoni S**, **Brehmer P**, **Dengler M**, **Dutra E**, **Faye S**, **Fischer E**, **van Heerwaarden C**, **Hohenegger C**, **Jarvinen H**, **Jochum M**, **Jung T**, **Jungclaus JH**, **Keenlyside NS**, **Klocke D**, **Konow H**, **Klose M**, **Malinowski S**, **Martius O**, **Mauritsen T**, **Mellado JP**, **Mieslinger T**, **Mohino E**, **Pawlowska H**, **Peters-von Gehlen K**, **Sarre A**, **Sobhani P**, **Stier P**, **Tuppi L**, **Vidale PL**, **Sandu I and Stevens B** (2025) nextGEMS: entering the era of kilometer-scale Earth system modeling, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-509.

Seifert A. and Beheng KD (2006) A two-moment cloud microphysics parameterization for mixed-phase clouds. part 1: model description. *Meteorology and Atmospheric Physics 92*(1), 45–66. https://doi.org/10.1007/s00703-005-0112-4.

Seifert A and Siewert C (2024) An ml-based p3-like multimodal two-moment ice microphysics in the icon model. https://doi.org/10.22541/essoar.170689097.75412140/v1.

Sharma S and Greenberg D (2024) Superdropnet: a stable and accurate machine learning proxy for droplet-based cloud microphysics [arXiv:2402.18354 [physics]]. https://doi.org/10.48550/arXiv.2402.18354.

Stevens B, **Acquistapace C**, **Hansen A**, **Heinze R**, **Klinger C**, **Klocke D**, **Rybka H**, **Schubotz W**, **Windmiller J**, **Adamidis P**, **Arka I**, **Barlakas V**, **Biercamp J**, **Brueck M**, **Brune S**, **Buehler SA**, **Burkhardt U**, **Cioni G**, **Costa-Surós M**, **Crewell S**, **Crüger T**, **Deneke H**, **Friederichs P**, **Henken CC**, **Hohenegger C**, **Jacob M**, **Jakub F**, **Kalthoff N**, **Köhler M**, **van Laar TW**, **Li P**, **Löhnert U**, **Macke A**, **Madenach N**, **Mayer B**, **Nam C**, **Naumann AK**, **Peters K**, **Poll S**, **Quaas J**, **Röber N**, **Rochetin N**, **Scheck L**, **Schemann V**, **Schnitt S**, **Seifert A**, **Senf F**, **Shapkalijevski M**, **Simmer C**, **Singh S**, **Sourdeval O**, **Spickermann D**, **Strandgren J**, **Tessiot O**, **Vercauteren N**, **Vial J**, **Voigt A and Zängl G** (2020) The added value of large-eddy and storm-resolving models for simulating clouds and precipitation. *Journal of the Meteorological Society of Japan Ser II 98*(2), 395–435. https://doi.org/10.2151/jmsj.2020-021.

Weiss P and Stier P (2024) Simulating the Earth system with interactive aerosols at the kilometer scale. https://doi.org/10.5194/egusphere-egu24-2359.

Yu S, **Hu Z**, **Subramaniam A**, **Hannah W**, **Peng L**, **Lin J**, **Bhouri MA**, **Gupta R**, **Lütjens B**, **Will JC**, **Behrens G**, **Busecke JJM**, **Loose N**, **Stern CI**, **Beucler T**, **Harrop B**, **Heuer H**, **Hillman BR**, **Jenney A**, **Liu N**, **White A**, **Zheng T**, **Kuang Z**, **Ahmed F**, **Barnes E**, **Brenowitz ND**, **Bretherton C**, **Eyring V**, **Ferretti S**, **Lutsko N**, **Gentine P**, **Mandt S**, **Neelin JD**, **Yu R**, **Zanna L**, **Urban N**, **Yuval J**, **Abernathey R**, **Baldi P**, **Chuang W**, **Huang Y**, **Iglesias-Suarez F**, **Jantre S**, **Ma P-L**, **Shamekh S**, **Zhang G**

and Pritchard M (2024) ClimSim-Online: a large multi-scale dataset and framework for hybrid ML-physics climate emulation. https://doi.org/10.48550/arXiv.2306.08754.

Zängl G, **Reinert D**, **Rípodas P and Baldauf M** (2014) The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society 141*(687), 563–579. https://doi.org/10.1002/qj.2378.

A. Appendix

Figure A1 shows the correlation between the considered input and output variables. The figure shows that the input and output features show low correlation.

Figure A2 illustrates the regression results individually for the output variables tendency of temperature, water vapor mmr, and cloud water mmr and their corresponding Shapley plots.

Figure A3 illustrates the regression results individually for the output variables tendency of rain mmr, snow mmr, and graupel mmr and their corresponding Shapley plots.

Figure A4 and A5 present an illustration of the performance of the combined ML model.

Figure A6 shows the corresponding frequency distributions of the microphysical tendencies (see Section 3.3).

Table A1 lists the variables considered in this work.

A.1. Computing resources

We are only able to run the simulation for this short amount of time because of the substantial computing resources necessary for running such a simulation (12 k node hours on the Levante cluster of the DKRZ for 30 days simulation) and for storing such a large volume of data (25 TB for the 20 days of high-resolution data when reduced to 3-hourly output).

A.2. Frequency distributions of microphysical tendencies

Histograms were generated to visualize the frequency distributions of the microphysical tendencies for the coarse-resolution simulation, the high-resolution ground truth, and the ML predictions (Figure A6). All distributions of the microphysical tendencies show a large peak of near-zero values. For this reason, a logarithmic scaling of the y-axis has been chosen to better visualize all relevant tendencies. The ML model (red) fails reproduce the tails of the tendency distributions of rain, snow, graupel, and cloud ice. The tendency distribution of temperature is well captured by the ML model, and the tails of the water vapor and cloud liquid water distributions are overestimated compared to the ground truth (gray). Even though the ML prediction does not capture the tails of these distributions very well, the model performs well for values near zero, which explains its high overall performance for these variables. In contrast, the coarse-resolution simulation (green) overestimates the frequencies of near-zero tendencies compared to the ground truth for most variables, except for cloud water. The ML predictions generally underestimate near-zero values, except in the case of cloud water, where an overestimation occurs. The histograms demonstrate reasonable similarities between the ML predictions and the ground truth in terms of the general shape of the distributions, confirming the good agreement of the zonal mean tendencies, as shown in Figures 5, A4 and A5.
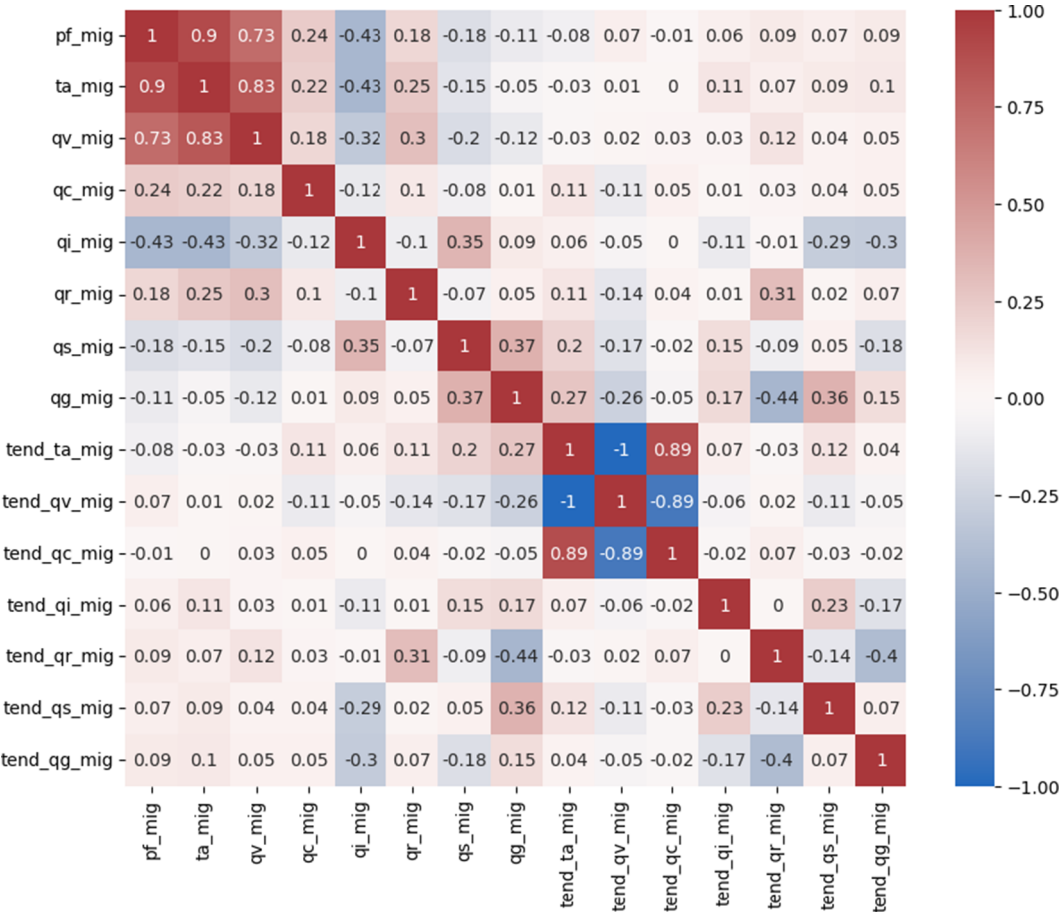
Figure A1. *Pearson autocorrelation matrix for all raw input and output parameters. For an explanation of the variable short names, the reader is referred to* Table A1. *The color bar ranges from negative correlation (blue) over no correlation (white) to positive correlation (red).*
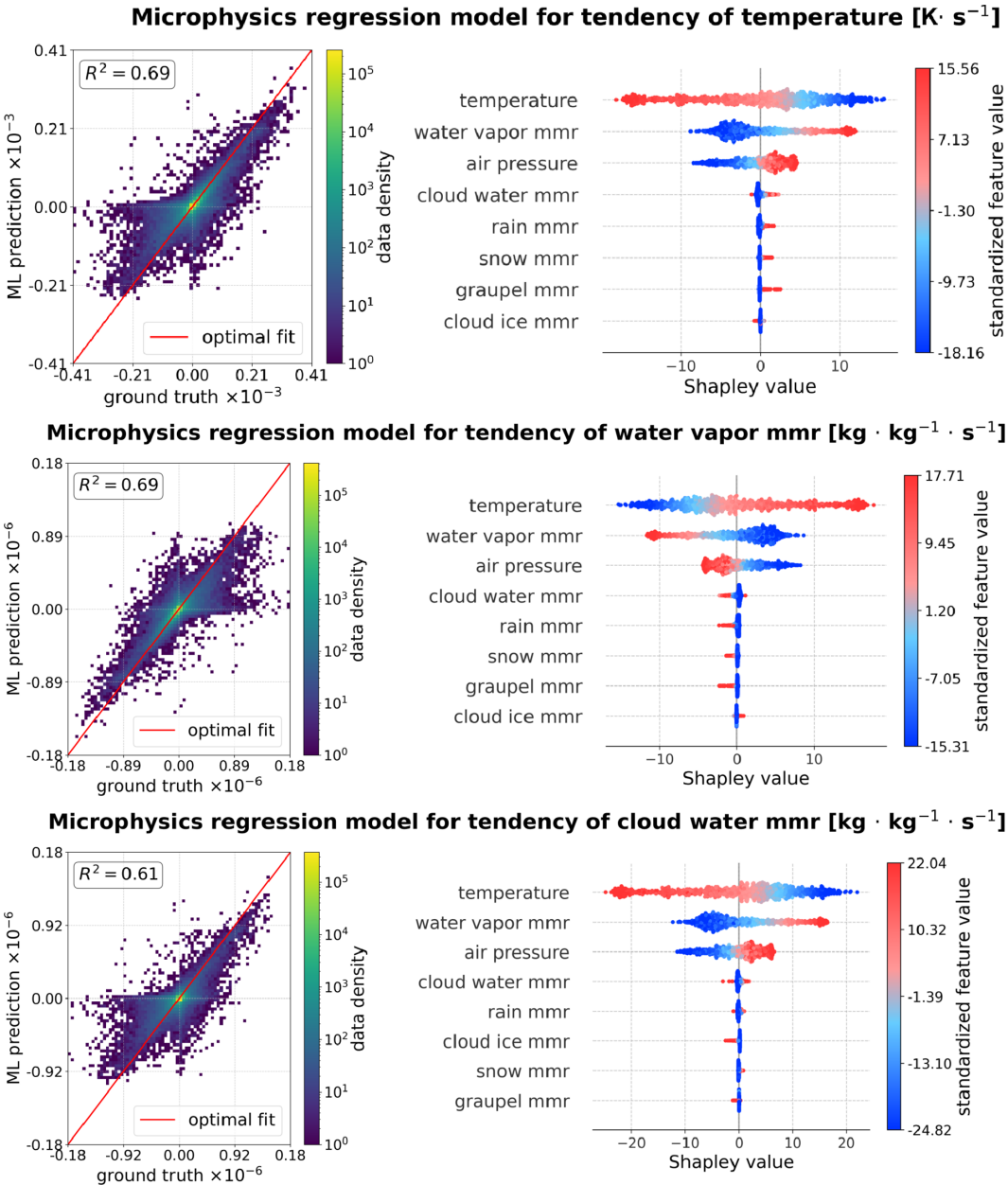
Figure A2. *Regression performance. The left images show the ML predictions of the regression model versus ground truth of the tendency of temperature, water vapor mmr, and cloud water mmr. The colors illustrate the density of the data on a logarithmic scale. The right images show the Shapley analysis for the tendency of temperature, water vapor mmr, and cloud water mmr. The input variables are listed in importance for the respective output variable from top to bottom. For each feature, red represents large values of a variable, and blue represents small ones. The x-axis shows the Shapley values.*
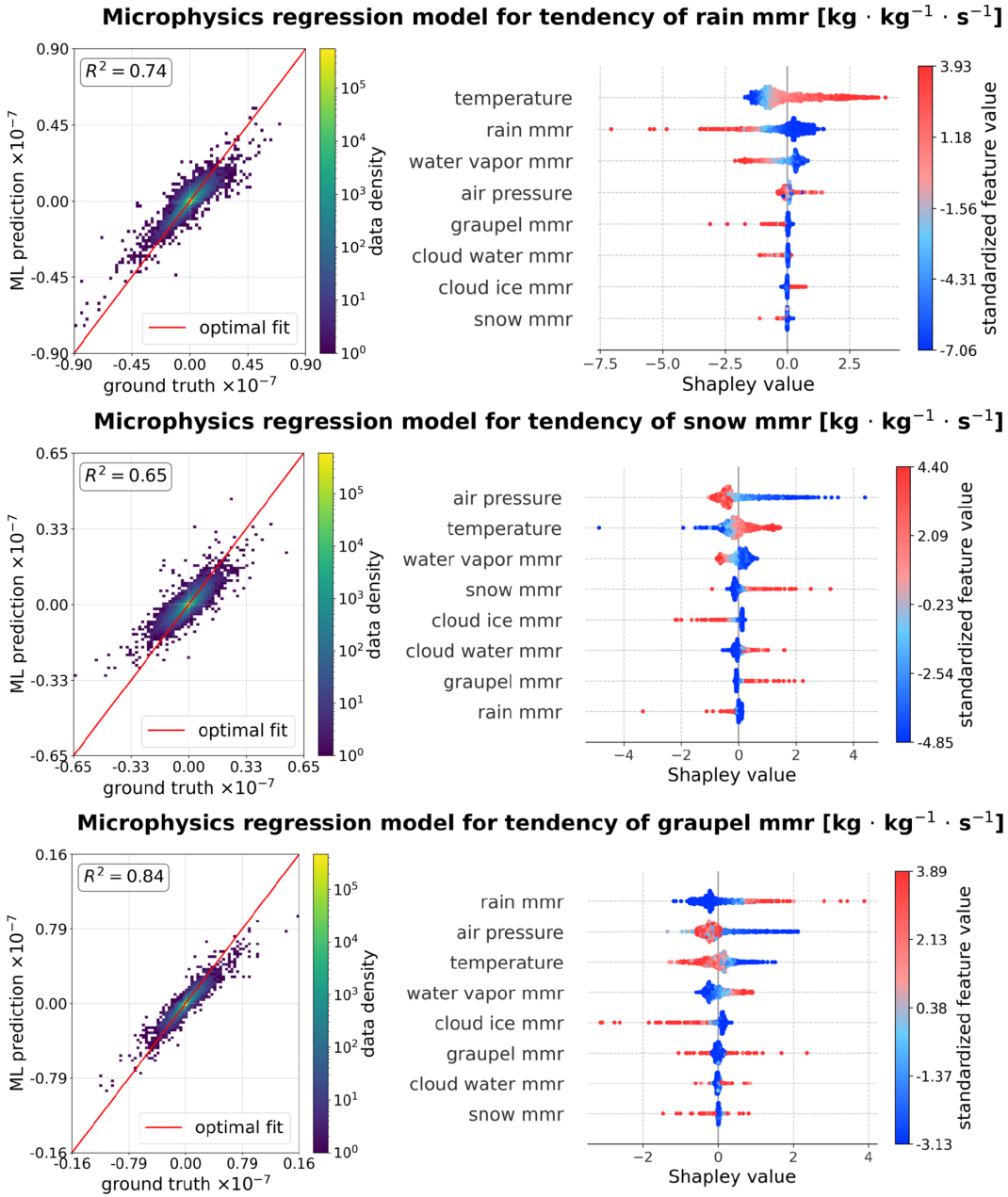
Figure A3. *Regression performance. The left images show the ML predictions of the regression model* versus *ground truth of the tendency of rain mmr, snow mmr, and graupel mmr. The colors illustrate the density of the data on a logarithmic scale. The right images show the Shapley analysis for the tendency of rain mmr, snow mmr, and graupel mmr. The input variables are listed in importance for the respective output variable from top to bottom. For each feature, red represents large values of a variable, and blue represents small ones. The x-axis shows the Shapley values.*
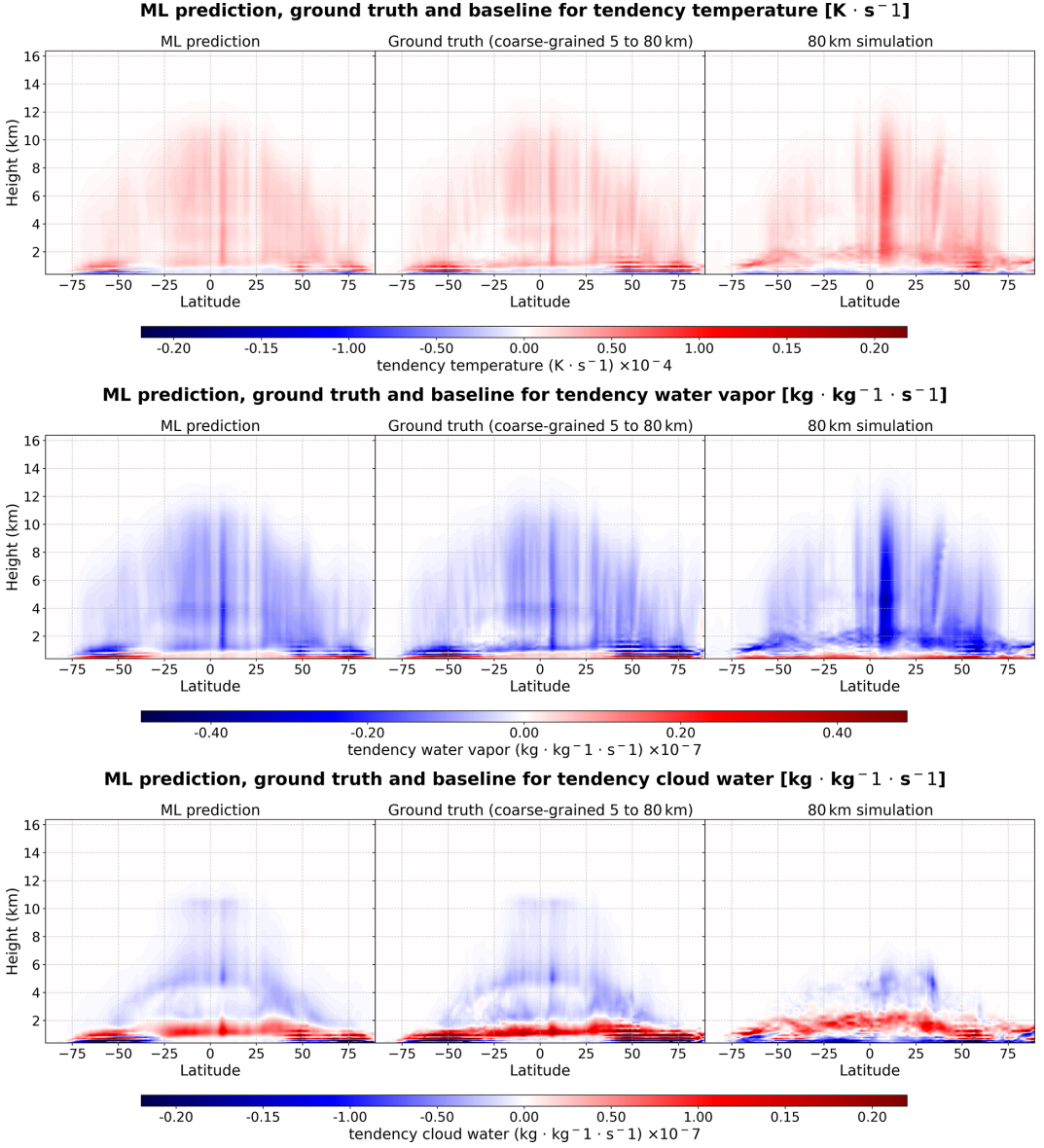
Figure A4. *ML prediction (left) of the tendency of temperature, water vapor, and cloud water mmr averaged over 3 validation days (24 timesteps in total) compared with the ground truth (center) and a coarse reference simulation (right) for a random day within the validation period. Colors represent the magnitude of the mmr tendency, averaged over longitude. The x-axis is latitude, the y-axis is height, and white areas indicate no change due to microphysics.*

ML prediction, ground truth and baseline for tendency cloud ice [kg · kg⁻1 · s⁻1]

ML prediction, ground truth and baseline for tendency rain [kg · kg⁻1 · s⁻1]

ML prediction, ground truth and baseline for tendency snow [kg · kg⁻1 · s⁻1]

Figure A5. *ML prediction (left) of the tendency of cloud ice, rain, and snow mmr averaged over 3 validation days (24 timesteps in total) compared with the ground truth (center) and a coarse reference simulation (right) for a random day within the validation period. Colors represent the magnitude of the mmr tendency, averaged over longitude. The x-axis is latitude, the y-axis is height, and white areas indicate no change due to microphysics.*
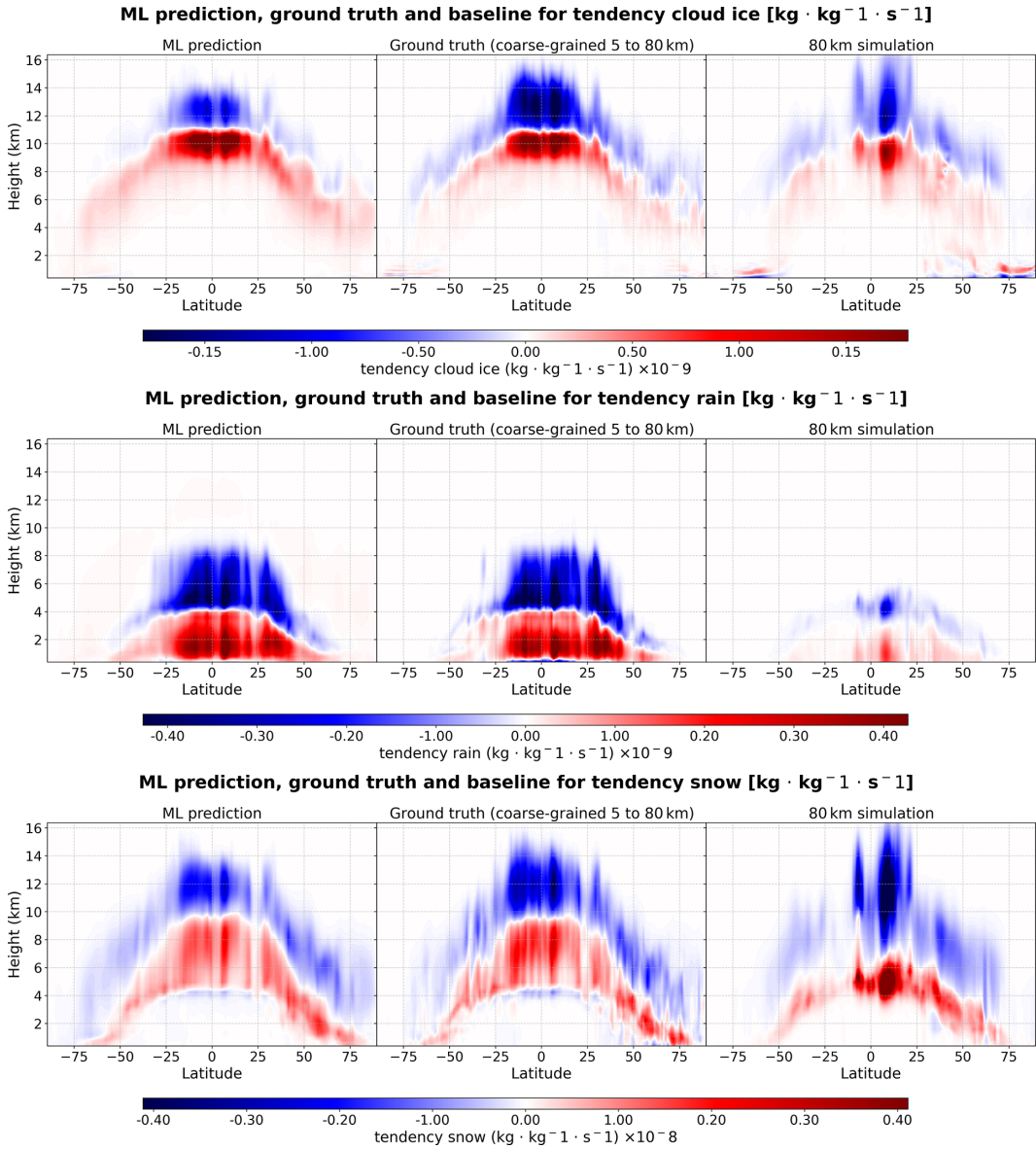
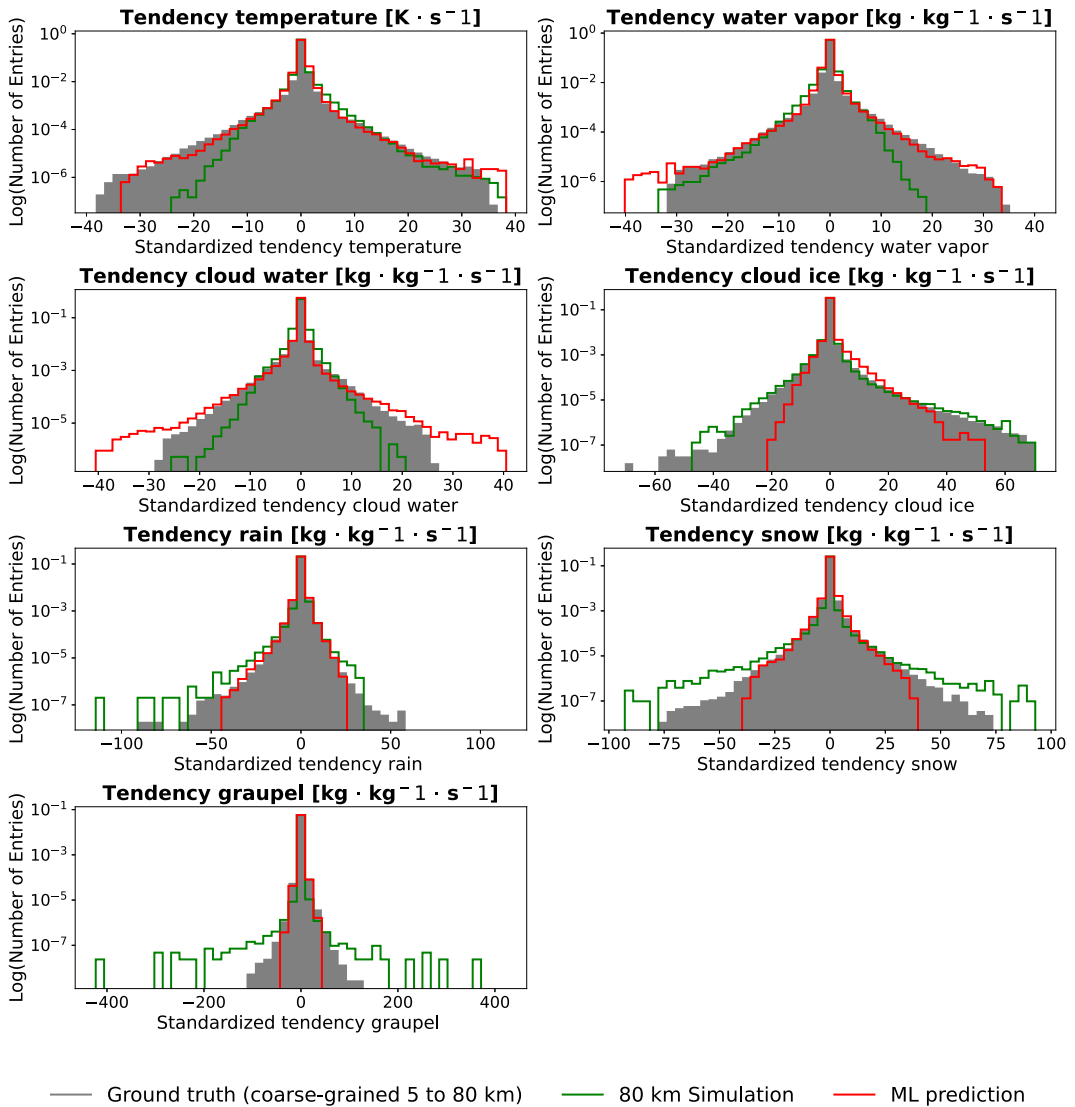**Distributions of microphysical tendencies (standardized)**

Figure A6. *Histograms of the microphysical tendencies averaged over 3 validation days (24 timesteps in total) for the high-resolution ground truth (gray), coarse-resolution simulation (green), and ML predictions (red). The histograms are standardized, and the number of entries is shown on a logarithmic scale to improve readability.*

Table A1. *Overview of all considered input and output features of the MLP model*

Type	Short name	Description	Unit
Input	pf_mig (p)	Air pressure	Pa
Input	ta_mig (T)	Temperature	K
Input	qv_mig (m_v)	Water vapor mmr	kg/kg
Input	qc_mig (m_c)	Cloud water mmr	kg/kg
Input	qi_mig (m_i)	Cloud ice mmr	kg/kg
Input	qr_mig (m_r)	Rain mmr	kg/kg
Input	qs_mig (m_s)	Snow mmr	kg/kg
Input	qg_mig (m_g)	Graupel mmr	kg/kg
Output	tend_ta_mig (ΔT)	Tendency of temperature	K/s
Output	tend_qv_mig (Δm_v)	Tendency of water vapor mmr	kg/(kg·s)
Output	tend_qc_mig (Δm_c)	Tendency of cloud water mmr	kg/(kg·s)
Output	tend_qi_mig (Δm_i)	Tendency of cloud ice mmr	kg/(kg·s)
Output	tend_qr_mig (Δm_r)	Tendency of rain mmr	kg/(kg·s)
Output	tend_qs_mig (Δm_s)	Tendency of snow mmr	kg/(kg·s)
Output	tend_qg_mig (Δm_g)	Tendency of graupel mmr	kg/(kg·s)