CAMBRIDGE
UNIVERSITY PRESS

APPLICATION PAPER

Cloudy with a chance of precision: satellite's autoconversion rates forecasting powered by machine learning

Maria Carolina Novitasari[1] , Johannes Quaas[2,3] and Miguel R. D. Rodrigues[1]

[1]Department of Electronic and Electrical Engineering, University College London, London, UK
[2]Leipzig Institute for Meteorology, Universität Leipzig, Leipzig, Germany
[3]ScaDS.AI - Center for Scalable Data Analytics and AI, Leipzig, Germany
Corresponding author: Maria Carolina Novitasari; Email: maria.novitasari.20@ucl.ac.uk

Abstract

Precipitation is one of the most relevant weather and climate processes. Its formation rate is sensitive to perturbations such as by the interactions between aerosols, clouds, and precipitation. These interactions constitute one of the biggest uncertainties in determining the radiative forcing of climate change. High-resolution simulations such as the ICOsahedral non-hydrostatic large-eddy model (ICON-LEM) offer valuable insights into these interactions. However, due to exceptionally high computation costs, it can only be employed for a limited period and area. We address this challenge by developing new models powered by emerging machine learning approaches capable of forecasting autoconversion rates—the rate at which small droplets collide and coalesce becoming larger droplets—from satellite observations providing long-term global spatial coverage for more than two decades. In particular, our approach involves two phases: (1) we develop machine learning models which are capable of predicting autoconversion rates by leveraging high-resolution climate model data, (2) we repurpose our best machine learning model to predict auto-conversion rates directly from satellite observations. We compare the performance of our machine learning models against simulation data under several different conditions, showing from both visual and statistical inspections that our approaches are able to identify key features of the reference simulation data to a high degree. Additionally, the autoconversion rates obtained from the simulation output and satellite data (predicted) demonstrate statistical concordance. By efficiently predicting this, we advance our comprehension of one of the key processes in precipitation formation, crucial for understanding cloud responses to anthropogenic aerosols and, ultimately, climate change.

Impact Statement

The ability to predict autoconversion rates directly from satellite data holds significant potential in enhancing our understanding of how clouds respond to anthropogenic aerosols. We develop a new process, powered by machine learning techniques, leading up to models that are able to predict autoconversion rates directly from satellite data. We argue that our approach is far less computationally expensive than atmospheric simulation-based methods, it offers reasonably good outcomes, and it can therefore easily leverage the wide availability of massive satellite data collection with long-term data provision. This exemplifies how important machine learning is for comprehending how aerosols affect clouds and, ultimately, climate change.

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

1. Introduction

Climate change represents one of the most pressing challenges afflicting our societies, with governments worldwide, non-governmental organizations, and other stakeholders placing increasing emphasis on measures to successfully tackle and mitigate its adverse effects. A pressing scientific challenge relates to the need to understand aerosol-cloud interactions because these currently represent one of the major sources of uncertainty in determining the radiative forcing responsible for climate change projections (Intergovernmental Panel on Climate Change 2021—IPCC [2021]).

Aerosols can directly affect climate via scattering and absorbing solar radiation, but they also affect climate indirectly by altering the properties of clouds, which impacts Earth's energy budget. In particular, an increase in aerosol concentrations may result in a larger number of droplets with smaller sizes, which, in turn, results in a higher cloud albedo (radiative forcing due to aerosol-cloud interactions, RFaci—Twomey [1974]); simultaneously, a more uniform distribution of smaller droplets reduces precipitation efficiency and increases cloud lifetime (one important adjustment to RFaci—Albrecht [1989], Gryspeerdt et al. [2019], Bellouin et al. [2020]). Therefore, determining the magnitude of force caused by aerosol-cloud interactions has been particularly challenging due to the intricate and interdependent nature of these processes, which involve complex feedback mechanisms. This complexity contributes to considerable uncertainty in our understanding of the radiative forcing of climate change.

One common way to study such complex interactions within the Earth system—i.e., the interaction between aerosols, clouds, and precipitation—is through the utilization of climate models. These models numerically solve the differential equations describing the fluid dynamics of the atmosphere and ocean on a discrete grid, with multiple layers across altitudes. This discretization is necessary due to the complexity of the climate system and the limited amount of computing power available, making it impossible to explain all processes occurring on Earth simultaneously. Processes that are smaller than the grid-scale of the model cannot be captured or represented directly by the model (IPCC [2021]). Parameterization is therefore required to resolve this problem. Parameterization is basically a simplified description, involving empirical parameters, of each small process that enables them to be incorporated into the model.

Alternatively, recent improvements in computer power have made it possible to rely on high-resolution models with smaller grid cells to enhance the representation of small-scale processes in the atmosphere in a realistic, large-domain setting (e.g., Stevens et al. [2020]). While these models excel at capturing small-scale phenomena, their applicability is typically limited to specific spatial domains and short periods due to the considerable computational complexity involved. For instance, one of the high-resolution simulation models that can be used to simulate small-scale processes in the atmosphere is the ICOsahedral Non-hydrostatic Large-Eddy Model (ICON-LEM—Zängl et al. [2015], Dipankar et al. [2015], Heinze et al. [2017]). This simulation model requires approximately 13 hours across 300 computer nodes, incurring a cost of approximately EUR 4000 just to generate a single hour of climate data over Germany (Costa-Surós et al. [2020]). Therefore, given such a high cost, it is imperative to devise alternative approaches to capture complex interactions within the Earth's climate system.

Our overarching goal—in line with other works showing that data-driven approaches can have a revolutionary impact in a variety of domains (Rolnick et al. [2019], Reuther et al. [2019])—is to understand whether machine learning approach can be used to address this challenge: in particular, we focus on the prediction of autoconversion rates—the rate at which small droplets (cloud droplets) transform into larger ones (raindrops) through the collision and coalescence of cloud droplets—directly from satellite observations. This is one of the key processes of precipitation formation for liquid clouds, which in turn is key to better understanding cloud responses to anthropogenic aerosols (specifically the cloud lifetime effect—Albrecht [1989]).

In view of the fundamental need to understand autoconversion rates, there have also been various recent attempts to predict the autoconversion rates by using machine learning techniques, such as Seifert and Rasp [2020], Chiu et al. [2021], and Alfonso and Zamora [2021]. The study by Seifert and Rasp [2020] utilizes

neural networks to explore cloud microphysical parameterizations, focusing on warm-rain formation processes, such as autoconversion, accretion, and self-collection. Enhanced parameterizations for auto-conversion and accretion rates in warm rain clouds were introduced by Chiu et al. [2021], informed by machine learning and optimization techniques. These parameterizations are based on in situ cloud probe measurements obtained during the Atmospheric Radiation Measurement Program field campaign in the Azores. Furthermore, Alfonso and Zamora [2021] developed a machine-learned parameterization using a deep neural network. This neural network was trained using a dataset of autoconversion rates derived from solving the kinetic collection equation over a wide range of droplet concentrations and liquid water contents.

Our research distinguishes itself from prior studies by shifting its focus. While prior research has predominantly focused on estimating autoconversion rates within simulation data and in-situ measurements, our investigation revolves around the more complex objective of directly estimating autoconversion rates from satellite observations.

Our motivation derives from the fact that satellite data provide extensive geographic coverage and frequent observations, which therefore have the potential to enhance our understanding of autoconversion rates in clouds on a scale that simulation-based models, due to computational and cost constraints alluded to earlier, cannot achieve. However, directly developing a machine learning model capable of predicting autoconversion rates from satellite observations is not a trivial task. Therefore, we adopt an innovative approach involving two phases leading up to the development of such a prediction model: (1) first, we leverage high-resolution climate model data to develop machine learning models capable of predicting autoconversion rates, and (2) second, we repurpose our best machine learning model to predict auto-conversion rates directly from satellite observations.

Our paper is structured as follows: In Section 2, we delve into the methodology employed in our study, specifically discussing the datasets, machine learning models, procedures adopted in our study, and the evaluation of the machine learning models. In Section 3, we present a number of experimental results showcasing that our approach can successfully predict autoconversion rates directly from satellite data. Finally, in Section 4, we conclude our paper, also offering insights into potential future directions for further investigation.

2. Methodology

We now describe the approach we adopt to develop machine learning models that can be used to estimate autoconversion rates directly from satellite observations. In summary, our methodology encompasses two stages as follows:

- The first stage involves building a set of machine learning models that can be used to deliver the autoconversion rates given the cloud's effective radius. The training and testing of these models are performed by leveraging high-resolution climate model data. Please refer to Sub-subsection 2.3.1 for more details.
- The second stage involves repurposing the (best) machine learning model to predict autoconversion rates directly from satellite observations. Please refer to Sub-subsection 2.3.2 for more details.

Our methodology is also briefly portrayed in Figure 1, where the left-hand side emphasizes how we use climate science-based data derived from a high-resolution climate science model to generate a dataset to train, validate, and test our machine learning models (under stage 1), whereas the right-hand side illustrates the use of the best machine learning model to directly predict autoconversion rates from satellite data (under stage 2).

We next describe in detail the various datasets (Subsection 2.1), machine learning models (Subsection 2.2), learning procedures (Subsection 2.3), and the evaluation of the machine learning models (Subsection 2.4), underlying these two stages.
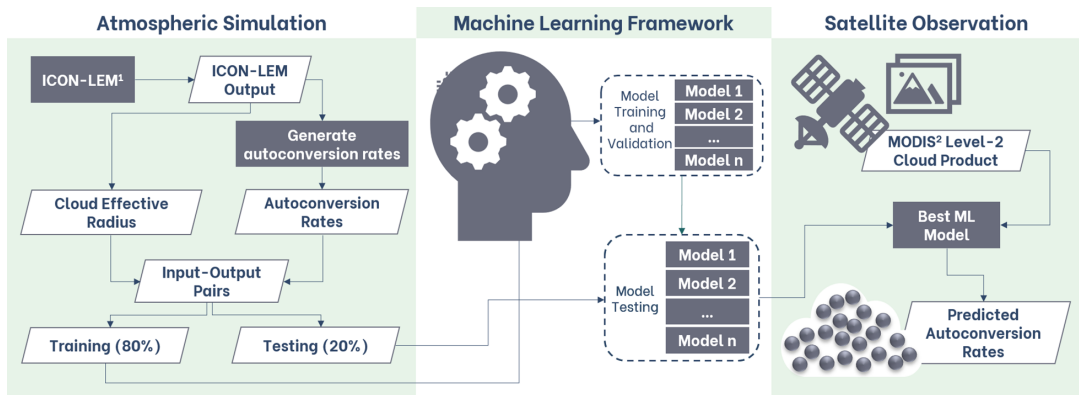
Figure 1. *The general framework of our proposed method. The left side of the image illustrates the climate science-based procedures we apply to our dataset to generate input-output pairs for training and testing. The center of the image represents our machine learning framework. The right side depicts the satellite observation data we used and the procedure to predict the autoconversion rates from the satellite data. 1 ICOsahedral Non-hydrostatic Large-Eddy Model; 2 Moderate Resolution Imaging Spectroradiometer.*

2.1. Data

2.1.1. Stage 1: climate model data

We first describe the data underlying stage 1 of our approach, where we develop some models capable of predicting autoconversion rates given the cloud effective radius, at each grid point in three-dimensional space, including latitude, longitude, and altitude or level. We rely on climate science-based procedures, leveraging data from high-resolution climate models, to generate datasets containing various input-output pairs consisting of cloud effective radius paired with the corresponding autoconversion rate.

In particular, we utilize data deriving from the ICON-LEM simulations over Germany on 2 May 2013 to train, validate, and test our machine learning models. The ICOsahedral non-hydrostatic model (ICON) is a unified model for global numerical weather prediction and climate simulations which is the result of a joint project between the German Meteorological Service (Deutscher Wetterdienst, DWD) and the Max Planck Institute for Meteorology (MPI-M) (Zängl et al. [2015]). The ICON Large Eddy Model (ICON-LEM) is an ICON version which can be run in a high-resolution large eddy mode. Detailed descriptions and comprehensive evaluations of ICON-LEM are provided in references Dipankar et al. [2015] and Heinze et al. [2017], respectively.

Specifically, we utilize data deriving from the ICON-LEM simulations relating to a time period approximately corresponding to the overpass times of the polar-orbiting satellites, specifically focusing on the time period of 09:55 UTC to 13:20 UTC. The period was chosen due to the fact that distinct cloud regimes occurred in the model domain on the chosen day (2 May 2013), providing an opportunity to investigate quite different elements of cloud formation and evolution (Heinze et al. [2017]).

The data covers 150 altitude levels in the vertical grid, with the grid reaching the top of the model at 21 km. The ICON-LEM simulations also cover 3 different horizontal resolutions, in two-way nests, namely 625, 312, and 156 m. Our study focuses on ICON-LEM with the highest resolution, 156 m, on the native ICON grid, and then regridded to a regular 1 km resolution to match the resolution of the MODIS data. This allows us to directly compare the simulated data with the observed data from MODIS.

Further, we also utilize data deriving from ICON numerical weather prediction (ICON-NWP) simulations (Kolzenburg et al. [2017], Haghighatnasab et al. [2022]) that were performed over the North Atlantic ocean for a week from 1 September 2014 to 7 September 2014 since these simulations capture completely different weather conditions, thereby offering an opportunity to further test the performance of our machine learning model. The ICON-NWP simulations for the Holuhraun volcano have 75 altitude

levels in the vertical grid, with the grid reaching the top of the model at 30 km, and a horizontal resolution of approximately 2.5 km.

The simulation data contains a number of variables associated with each grid point such as cloud effective radius, cloud optical thickness, etc. However, due to the vast amount of storage, the large-domain high-resolution simulation of Heinze et al. [2017] did not store all the quantities necessary for our study. Instead, state quantities and process rates had to be diagnosed (recomputed) from the available three-dimensional instantaneous output. In particular, the two-moment microphysical parameterization of Seifert and Beheng [2006] was used to derive the autoconversion rates, in our training, validation, and testing data.

The autoconversion parameterization given by Seifert and Beheng [2006]—which applies a double-moment bulk cloud and precipitation microphysical scheme with prognostic mass and number for different hydrometeor classes—is given by:

$$\frac{\partial L_r}{\partial t}\bigg|_{au} = \frac{k_{cc}}{20x^*} \frac{(v_c+2)(v_c+4)}{(v_c+1)^2} L_c^2 + \bar{x}_c^2 \left[1 + \frac{\Phi_{au}(\tau)}{(1-\tau)^2}\right] \frac{\rho_0}{\rho} \tag{2.1}$$

This equation—which describes the rate of change of rainwater content (specific mass) (L_r), due to the process of autoconversion—makes the assumption that the size distribution of cloud droplets follows a hyper-gamma distribution and the size distribution of raindrops follows an exponential distribution. The hyper-gamma distribution is a probability distribution that is characterized by a width parameter (v_c), whereas the exponential distribution is a probability distribution that is commonly used to model the size distribution of raindrops (Marshall and Palmer [1948]).

The variable τ captures the ratio of the cloud water content, L_c, to the sum of the cloud and rain water content, i.e., $\tau = 1 - (L_c/(L_c + L_r))$. The constant k_{cc}, which represents the cloud-cloud collision rate, has a value of $4.44 \times 10^9 \, \text{m}^3 \text{kg}^{-2} \text{s}^{-1}$, as determined by the study conducted by Pinsky et al. [2001]. The density of air at sea level, ρ_0, is assumed to be equal to $1.225 \, \text{kg m}^{-3}$. The varying air density is denoted as ρ. The function $\Phi_{au}(\tau)$ is defined as $400\tau^{0.7}(1 - \tau^{0.7})^3$. A drop mass of $x^* = 2.6 \times 10^{-10}$ kg is used as a threshold to distinguish between cloud droplets and raindrops. This value corresponds to an individual drop with a radius of 40 μm and a liquid water density of 1000kg m^{-3}. Additionally, $\bar{x}_c$ represents the mean mass of cloud droplets in kg.

The computation of the autoconversion rates utilizing the method outlined in the study by Seifert and Beheng [2006], requires several additional pieces of information. Specifically, we need the mass density of cloud droplets or the cloud water content (L_c), the mass density of raindrops (L_r), and the air density (ρ). Appendix A contains the information required to compute these values.

Equation 2.1 highlights that this autoconversion rate formulation functions as a highly non-linear expression based on specific cloud mass. In this context, specific cloud mass, cloud droplet number concentration, and cloud effective radius exhibit interconnectedness. The droplet number concentration within the simulations serves as a prognostic variable, originating from the activation of cloud condensation nuclei (CCN), prescribed to vertical variation but constant in the horizontal and in time. Several sink processes of droplet number include autoconversion itself, accretion, and also droplet evaporation following turbulent mixing.

Predicting the autoconversion rates holds significance. The "real" process is the coagulation of droplets and drops across the size spectrum. While categorizing the particle size spectrum into two distinct classes is a simplification, it remains a justified one for liquids. This approach proves particularly useful in understanding how clouds lose their water through precipitation.

Finally, we specifically extracted the autoconversion rates pertaining to cloud tops because this reflects satellite-like data. The cloud-top autoconversion rates are determined by selecting autoconversion rates at the level where, integrating from the cloud-top, the cloud optical thickness exceeds 1. This optical thickness threshold represents the penetration depth for retrieving cloud microphysical quantities by the optical satellite sensors. While our approach provides practicality, it might not directly translate to scattering scenarios, particularly regarding wavelengths relevant to MODIS optical properties retrieval

(Platnick [2000]). While acknowledging that this simplification might not perfectly align with scattering cases and could potentially have limitations, we believe it still fulfills its testing purpose. It is important to note that we exclusively utilize this quantity for testing purposes. Our training model remains unaffected, as each input-output variable is trained with its corresponding vertical variety.

In summary, by adopting this procedure, we are able to create a dataset containing a series of data points consisting of cloud effective radius—autoconversion rate pairs corresponding to different points on the grid.

2.1.2. Stage 2: satellite data

We now turn to the data underlying stage 2 of our approach where we repurpose the learned machine learning (best) model to predict the autoconversion rates directly from satellite observations.

Concretely, we utilize Collection 6 of level-2 (L2) cloud product of Terra and Aqua of the Moderate Resolution Imaging Spectroradiometer (MODIS) (Platnick et al. [2017], Platnick et al. [2018])—MOD06 and MYD06, respectively. The MODIS L2 cloud product contains information on both cloud-top properties and cloud optical properties.

Onboard the Terra and Aqua satellites, MODIS is one of the most commonly utilized remote sensing instruments for Earth science research. Terra's orbit crosses the equator from north to south in the morning, whereas Aqua's orbit crosses the equator from south to north in the afternoon. The approximate overpass time for the Terra satellite is around 10.30 hours, whereas, for the Aqua satellite, it is around 13.30 hours of local solar time.

Additionally, MODIS employs 36 spectral bands spanning from visible to infrared wavelengths. MODIS visible and near-infrared channel radiances are employed to determine cloud particle phases, effective cloud particle radii, and cloud optical thickness. We utilize the cloud effective radius at each grid point in a two-dimensional space, which corresponds to a specific range of latitude and longitude with a resolution of 1 km. The machine learning model ingests this cloud effective radius to deliver an estimate of the cloud-top autoconversion rates.

Regarding the MODIS data, we have selected cloud effective radius (CER) derived from the 2.1 µm channel. This decision was based on considering the following options: the 3.7 µm option resulted in an excessively small CER distribution, while the 1.6 µm option would yield an overly large distribution (Zhang and Platnick [2011]). Therefore, the selection of the middle value, 2.1 µm, seemed most suitable given these considerations.

2.1.3. Data preprocessing

In line with standard practices, we further preprocess the data (for both simulation output and satellite data) in order to train, validate, and test our various machine learning models. In particular, we adopt the following procedure:

Preprocessing step 1: First, we exclusively utilize data relating to cloudy grid columns that are further filtered by selecting cloud effective radius smaller than or equal to 40 µm, since this specific threshold of 40 µm is also used to separate cloud droplets and raindrops as used in Seifert and Beheng [2006]. The reason for adopting such a filtering process has to do with the fact that we intend to focus exclusively on cloud droplets, since our primary focus lies in the autoconversion process, rather than accretion or self-collection processes.

Preprocessing step 2: Second, we transform and normalize both input data (cloud effective radius) and output data (autoconversion rate) according to the transformation given by:

$$z' = \frac{\log_{10}(z) - \mu}{\sigma} \tag{2.2}$$

where z' denotes the normalized data point (input or output) and z denotes the original data point, μ corresponds to the mean value of the logarithmically transformed input or output data points, and σ corresponds to the standard deviation of the logarithmically transformed data points. It is worth noting that

this normalization operation is carried out independently for points from the training and testing sets to avoid contamination. The rationale for adopting this transformation/normalization procedure is to effectively deal with extremely small values of autoconversion rates and cloud effective radius (this is particularly important for the autoconversion rates since they exhibit minuscule values). Furthermore, by applying this transformation, it becomes more manageable to disentangle the relationship between these two variables and to guarantee that data features are not too far away from zero in order to stabilize the learning process.

Before proceeding with both preprocessing steps mentioned above, we initially proceeded to split the relevant datasets into subsets for training, validating, and testing the machine learning models. Notably, in stage 1 of our approach, we split the data into a training set (containing 80% of the original data) and a testing set (containing 20% of the original data). For the neural network models, we use 20% of the training set as the validation set, serving as a means to fine-tune the model's performance. Meanwhile, for the random forest model, we use 3-fold cross-validation. In stage 2 of our approach, we only use a testing set since we are only assessing the performance of the best model under stage 1 using real satellite data.

2.2. Machine learning models

We have selected, trained, and tested various machine learning models to predict desired output from input, including linear regression, polynomial regression, random forest (RF—Breiman [2001]), shallow neural network (NN), and deep neural network (DNN—Schmidhuber [2015]). We briefly describe some of these models in the sequel.

2.2.1. Random forest (RF)

A random forest is an ensemble method that is constructed from decision tree algorithms (Breiman [2001]). This technique uses bootstrap aggregation (bagging) to select a random sample from a training set where individual data points can be selected multiple times. The term "forest" refers to a group of trees, in this case, decision trees.

The bagging technique is applied to collect several sets of random samples. Each sample set is then dealt with by each individual decision tree. The final prediction delivered by a random forest is determined by averaging the predictions of each individual decision tree.

A random forest uses a range of hyperparameters, including N_t, min_s, min_l, and max_d. The hyperparameter N_t represents the number of trees in a random forest; the number of trees influences the complexity of the model. The hyperparameter min_s governs the minimum number of samples required to split a tree node in a random forest. In other words, if the node contains samples that exceed or equal to min_s, it can be further subdivided. A small value of min_s may result in a model that overfits the data; conversely, a large value of min_s may result in a model that underfits the data. The hyperparameter min_l indicates the least amount of samples which, notwithstanding a split, should appear in leaf nodes. This means that any split point within any level must leave at least min_l training samples on the left and right branches. The hyperparameter max_d defines the depth of the tree, which is the furthest distance between the root and leaf nodes.

These hyperparameters can be optimized by using techniques such as grid search and random search to balance the trade-off between model performance and complexity. Since the performance of an RF is strongly influenced by its hyperparameters, we conduct a random search to find reasonable hyperparameter values. Our final hyperparameters are shown in Table 1.

2.2.2. Neural networks

A neural network is a type of machine learning model inspired by the work of the human brain, simulating how neurons interact with each other, which was first proposed by McCulloch and Pitts [1943]. A neural network is made up of several layers, including an input layer (ingesting the input data), one or more hidden layers, and an output layer (delivering the output data). Each individual layer is composed of

Table 1. *Random forest hyperparameters selected in this study*

Variable	Value
N_t	100
min_s	15
min_l	8
max_d	10

various neurons, connecting to other neurons in the previous layer, and connecting to neurons in the subsequent layer, with each connection carrying a corresponding weight.

The Deep Neural Network (DNN—Schmidhuber [2015]) is a neural network with multiple hidden layers. Let $x_j^{(l)}$ be the value of the j-th neuron in layer l, $x_i^{(l-1)}$ be the value of the i-th neuron in layer $l-1$, $w_{ij}^{(l)}$ be the weight of the connection between neuron i in layer $l-1$ and neuron j in layer l, and $b_j^{(l)}$ be the bias term of neuron j in layer l. Then, we can express the value of $z_j^{(l)}$ (before applying the activation function) in terms of the values of $x_i^{(l-1)}$, $i = 1, \ldots, n^{(l-1)}$, where $n^{(l-1)}$ is the number of neurons in the layer $l-1$, as follows:

$$z_j^{(l)} = \sum_{i=1}^{n^{(l-1)}} w_{ij}^{(l)} x_i^{(l-1)} + b_j^{(l)} \tag{2.3}$$

The output of this neuron is then passed through the activation function g:

$$x_j^{(l)} = g\left(z_j^{(l)}\right) \tag{2.4}$$

where $g(\cdot)$ is a nonlinear activation function such as a rectified linear unit (ReLU). The input to the neural network (at the very first layer) corresponds to the various data features and the output of the neural network (at the very final layer) corresponds to the desired prediction.

A DNN is parameterized by various parameters, i.e., the weights and the biases corresponding to the different neural network layers. The parameters are adjusted during training using the backpropagation algorithm to minimize a desired loss function measuring the discrepancy between the predicted output and the reference simulation data, leveraging the availability of a training set.

We developed two neural network models—one very simple model (shallow NN) and another slightly more complex model (pyramid-shaped DNN) to evaluate the impact of alterations to the architecture of neural networks. The first model, referred to as the shallow NN or NN, is a basic neural network that has only one hidden layer consisting of 64 neurons. In contrast, the second model, referred to as the pyramid-shaped DNN or DNN, employs a more complex architecture that encourages hierarchical learning in the network. However, it is worth noting that this architecture was chosen arbitrarily and alternative architectures can also be employed. The DNN architecture of our machine learning model, with cloud effective radius (r_e) as input and autoconversion rates (a_u) as output, is depicted in Figure 2. The model contains five fully connected hidden layers, with node counts increasing by a factor of two for each subsequent layer—64 nodes in the first layer, 128 nodes in the second layer, and so on until the fifth layer.

Both the DNN and NN models use Leaky ReLU as the activation function at every hidden layer. As for the loss function for both models, we employed mean squared error (MSE). We choose MSE as our primary metric for several reasons. MSE's inherent property of squaring errors provides greater weight to larger deviations, which proves advantageous in situations where penalizing larger deviations is crucial, enhancing the model's sensitivity to outliers. Additionally, its smoothness and differentiability across its entire domain render it well-suited for optimization algorithms, simplifying gradient computations and thus streamlining the training process. Moreover, MSE's representation as an average of squared differences between predicted and actual values contributes to its interpretability by directly measuring
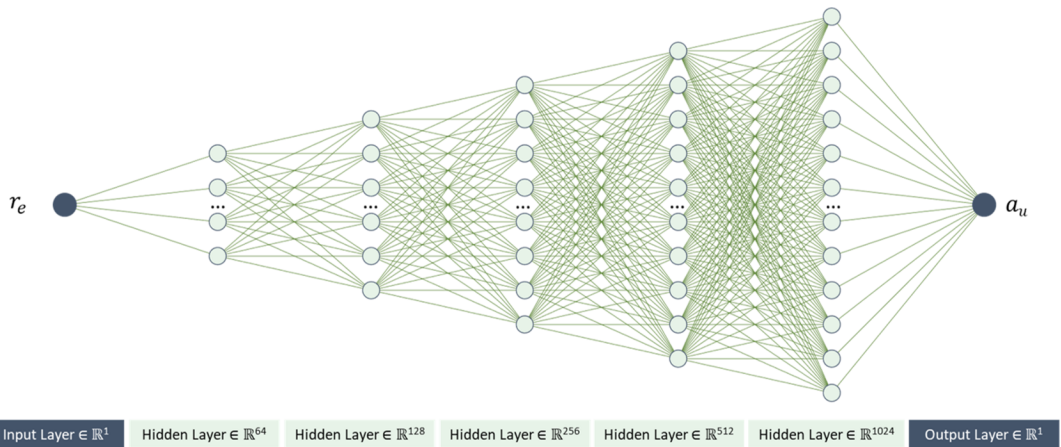
Figure 2. *DNN architecture of our machine learning model.*

the average magnitude of errors. Finally, MSE stands as the most common metric in regression problems, aligning with our objective of minimizing disparities between predicted and actual continuous numerical values, in this case, autoconversion rates. Additional experiments involving different loss functions are also included in Appendix B.1.

To optimize the models, we employed Adam's optimizer and utilized the KerasTuner package (O'Malley et al. [2019]) to determine the optimal number of training samples per batch. KerasTuner is a python package to solves the challenge of hyperparameter search through its user-friendly and scalable hyperparameter optimization framework.

2.3. Training/testing procedure

We next describe further the settings underlying our training and/or testing procedures relating to stages 1 and 2.

2.3.1. Stage 1—predicting the autoconversion Rates from atmospheric simulation model output

As discussed earlier, this phase involves training machine learning models which are able to predict autoconversion rates given the cloud effective radius, at each grid point in three-dimensional space, including latitude, longitude, and altitude or level, by leveraging a dataset containing input-output pairs derived from the atmospheric simulations (as described earlier).

In particular, as described in the previous subsections, we train various machine learning models of different complexities, including linear regression, polynomial regression, random forest, and neural networks, to determine their effectiveness. These models are trained using the dataset derived from the ICON-LEM simulations over Germany, as explained earlier.

Subsequently, we evaluate our best model using different testing datasets and scenarios associated with the ICON-LEM simulations over Germany and the ICON-NWP simulations over Holuhraun, as follows:

1. *ICON-LEM Germany*: In this testing scenario, we evaluate the performance of our machine learning models using the same data that was utilized during its training process. Once again, this data, which consists of a set of cloud effective radius and autoconversion rates corresponding to different points in three-dimensional space, was collected through the use of ICON-LEM simulations specifically over Germany. The testing data, however, differs from the training data as we focus on a different time period, specifically 2 May 2013 at 1:20 pm. This approach enables us to assess the model's generalization capability to new data within the same region and day, while considering significant weather variations that evolved considerably (Heinze et al. [2017]).

2. *Cloud-top ICON-LEM Germany*: In this testing scenario, we evaluate the performance of our machine learning model by utilizing the same data as in the previous scenario, with the exception that we are only considering the cloud-top information of the data. The data involved in this testing scenario is a collection of cloud-top autoconversion rates and cloud effective radius pairs associated with 2D spatial points, representing a specific range of latitude and longitude, as well as a specific altitude for each point, specifically the cloud-top height for each point. We extract this cloud-top 2D data from the 3D atmospheric simulation model by selecting the variable value at any given latitude and longitude where the cloud optical thickness exceeds 1, integrating vertically from the cloud-top.

3. *Coarser Horizontal Resolution of Cloud-top ICON-LEM Germany*: This testing scenario is designed to evaluate the performance of our machine learning model by utilizing the same data, but at different resolutions. Specifically, we have modified the resolution of cloud-top ICON-LEM Germany from 1 km to 2.5 km and 10 km. The evaluation of a model's performance in the presence of data with different spatial resolutions is important since it can help us understand the model's ability to make accurate predictions in real-world applications where data with varying spatial resolutions may be encountered.

4. *Cloud-top ICON-NWP Holuhraun*: This final testing scenario uses distinct data from that of previous scenarios. In particular, we use the cloud-top of ICON-NWP Holuhraun data that was acquired at a different location, time, and resolution compared with the data used in the previous scenarios. The ability of the model to perform well in the presence of new data is important in many practical applications, allowing the model to make accurate predictions on unseen data, adapt to varying geographical locations, and adapt to different meteorological conditions.

Overall, these testing scenarios are crucial in evaluating the performance and robustness of our machine learning models and can provide valuable insights that can be used to improve the model's accuracy and usefulness in a wide range of applications.

2.3.2. Stage 2—Predicting the autoconversion rates from satellite data

Finally, as discussed earlier, this stage focuses on predicting the autoconversion rates directly from satellite data. Notably, we do not train new machine learning models but instead simply select the most promising model derived during the previous stage to forecast the autoconversion rates of the satellite data-derived inputs, specifically the cloud effective radius. The best machine learning model is then tested on several specific scenarios, which can be found in Table 2.

The reasons behind our selection of those specific datasets at those particular times and locations are twofold. Firstly, the chosen datasets contained substantial cloud data, contributing significantly to our

Table 2. *Scenarios to test our NN model on satellite data: MODIS data corresponding to ICON-LEM and ICON-NWP simulations with specific time and area*

No	Correspond to	Date	Time	Lat (°N)	Lon (°E)
1	ICON-LEM Germany	2 May 2013	11:40 am	47.50° to 54.50°	5.87° to 10.00°
2	ICON-LEM Germany	2 May 2013	1:20 pm	47.50° to 54.50°	5.87° to 10.00°
3	ICON-NWP Holuhraun	1 September 2014	7 am	72° to 77°	−13° to 20°
4	ICON-NWP Holuhraun	2 September 2014	11 am	73° to 80°	−60° to −11°
5	ICON-NWP Holuhraun	3 September 2014	10 am	64° to 77°	10° to 20°
6	ICON-NWP Holuhraun	4 September 2014	2 pm	60° to 75°	−40° to 0°
7	ICON-NWP Holuhraun	5 September 2014	1 pm	50° to 58° N	−19° to 14°
8	ICON-NWP Holuhraun	6 September 2014	5 pm	50° to 54° N	−60° to −52°
9	ICON-NWP Holuhraun	7 September 2014	8 am	72° to 80°	−15° to 20°

study. Secondly, the data available at those specific times and locations were present in both the MODIS and ICON datasets, offering valuable consistency and comparability for our analysis. The use of these specific datasets allows us to evaluate the performance of the selected model under various geographical conditions as well as different time periods.

2.4. Model evaluation

The performance of each machine learning model is evaluated by calculating various metrics—such as R^2, MAPE (mean absolute percentage error), RMSPE (root mean squared percentage error), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM)—on the testing data. However, given that the autoconversion rates are presented in very small values, we have carried out a transformation of the reference simulation autoconversion rates and the predicted ones prior to the calculation of these various error metrics, involving applying a base-10 logarithm to these quantities followed by normalization to the range 0 to 1. These transformations have also enabled us to offer better visualizations of the relevant quantities on a two-dimensional grid, allowing a meaningful comparison between the predictions and the reference simulation data.

The R^2 metric is used to find out how well the regression model describes the relationship between the predicted values and actual values. In addition, the MAPE and RMSPE metrics are expressed as percentages making it easy to convey the performance of a model to individuals that may not possess familiarity with the particular data. MAPE is a metric that measures the average percentage difference between the predicted and actual values of a variable. On the other hand, RMSPE is similar to MAPE but it also takes into account not only the absolute difference between predicted and actual values but also their squared difference, which can be advantageous when identifying larger errors is crucial. Both metrics are commonly used in forecasting as they provide a clear indication of the percentage error in the model's predictions.

Finally, in view of the fact that we report some results using images, we also use two other measures that are widely adopted to evaluate image quality. In particular, we use PSNR and SSIM to measure the similarity of a predicted image with the reference simulation image. PSNR reports the ratio between the maximum signal value and the noise present in the image in decibel (dB). A higher PSNR value indicates superior image quality. SSIM, on the other hand, evaluates the structural similarity between the predicted and reference simulation images, taking into account attributes such as luminance, contrast, and structure. The resulting index ranges from -1 to 1, with a value of 1 indicating a perfect match between the images.

3. Experimental results

We now report our experimental results relating to the prediction of autoconversion rates from the atmospheric simulation model output and satellite data, with a variety of machine learning models and scenarios that have been developed using the procedure described earlier.

3.1. Autoconversion on simulation model (ICON)

We recall that in this scenario we have trained a variety of machine learning models, including simple and more complex ones, that are able to predict autoconversion rates given cloud effective radius; we have then tested the best machine learning model under a variety of scenarios that have been outlined in Subsection 2.3. The main observations relating to the variety of machine learning models and various testing scenarios are as follows:

ICON-LEM Germany: Table 3 showcases a variety of evaluation metrics pertaining to the prediction of autoconversion rates through different machine learning models in this testing scenario. The results indicate that simple linear regression produces reasonably good outcomes, with R^2 and SSIM values exceeding 88%. Furthermore, employing second-order polynomial regression leads to a slight but noticeable improvement in outcomes.

Table 3. *Evaluation of the autoconversion rates prediction results on the ICON-LEM simulation model using various machine learning models—linear regression (LR), second-order polynomial regression (PR), random forest (RF), shallow neural network (NN), and deep neural network (DNN)—over Germany testing scenario (scenario 1)*

Method	R^2	MAPE	RMSPE	SSIM	PSNR (dB)
LR	88.77%	10.86%	15.28%	88.39%	25.56
PR	90.03%	9.57%	12.22%	89.96%	26.08
RF	90.53%	9.15%	11.54%	90.09%	26.30
NN	90.54%	9.15%	11.35%	90.11%	26.31
DNN	90.54%	9.15%	11.29%	90.11%	26.31

Among all the models, the best results tend to be delivered by DNN, with R^2 and SSIM surpassing 90%, however, NN and random forest also perform well. The shallow NN and DNN yield comparable results, though the RMSPE values for the shallow NN are slightly higher than those for the DNN.

Figure 3 displays the plot of predicted autoconversion rates against the reference simulation autoconversion rates for the various machine learning models for this testing scenario, in order to illustrate further performance differences. We can see that second-order polynomial regression outperforms linear regression, whereas random forest produces results that are essentially identical to the NN and DNN models. We can also see that employing either of these strategies yields fairly decent results that can be applied to the atmospheric simulation model. Our findings also suggest that there is a strong association between the cloud effective radius or liquid cloud particle size and the autoconversion rates.

It is worth noting that the differences in performance among some models, especially RF, NN, and DNN, do not exhibit significance. This leads us to hypothesize that altering the architecture of the ML model using the same dataset may not significantly enhance performance. This hypothesis could be explored further by incorporating uncertainty estimation as a potential avenue for future research.

Ultimately, despite the fact that the DNN model produced slightly better results—albeit the difference not being significant—than competing models, we deem the NN model to be superior due to its ability to produce comparable outcomes with significantly fewer trainable weights. Therefore, we further showcase the NN performance on the ICON-LEM Germany and ICON-NWP Holuhraun, encompassing different types of settings.

Finally, given our ability to predict autoconversion rates from cloud effective radius with reasonable accuracy using linear regression (LR) and polynomial regression (PR) models, as shown in this scenario, we can also immediately write the respective input–output relationship in order to shed some light on the interplay between the autoconversion rates and cloud effective radius. This not only allows for simple calculations but also provides interpretability.

In order to make the calculations more accurate, we use the logarithmic base 10 form of the cloud effective radius. Furthermore, to standardize the data, we also use the logarithmic mean (the mean of the logarithm), $m_{\mathrm{e}}^{\log}$, and the logarithmic standard deviation (the standard deviation of the logarithm), $s_{\mathrm{e}}^{\log}$, of the cloud effective radius. We normalize r_{e} (cloud effective radius) using the following equation:

$$\hat{r}_{\mathrm{e}}^{\log} = \frac{\log_{10}(r_e) - m_{\mathrm{e}}^{\log}}{s_{\mathrm{e}}^{\log}}.$$

Following the application of the simple linear regression formulation $y = a + bx$, we obtain the coefficients a and b after training the linear regression model, with x and y serving as the respective
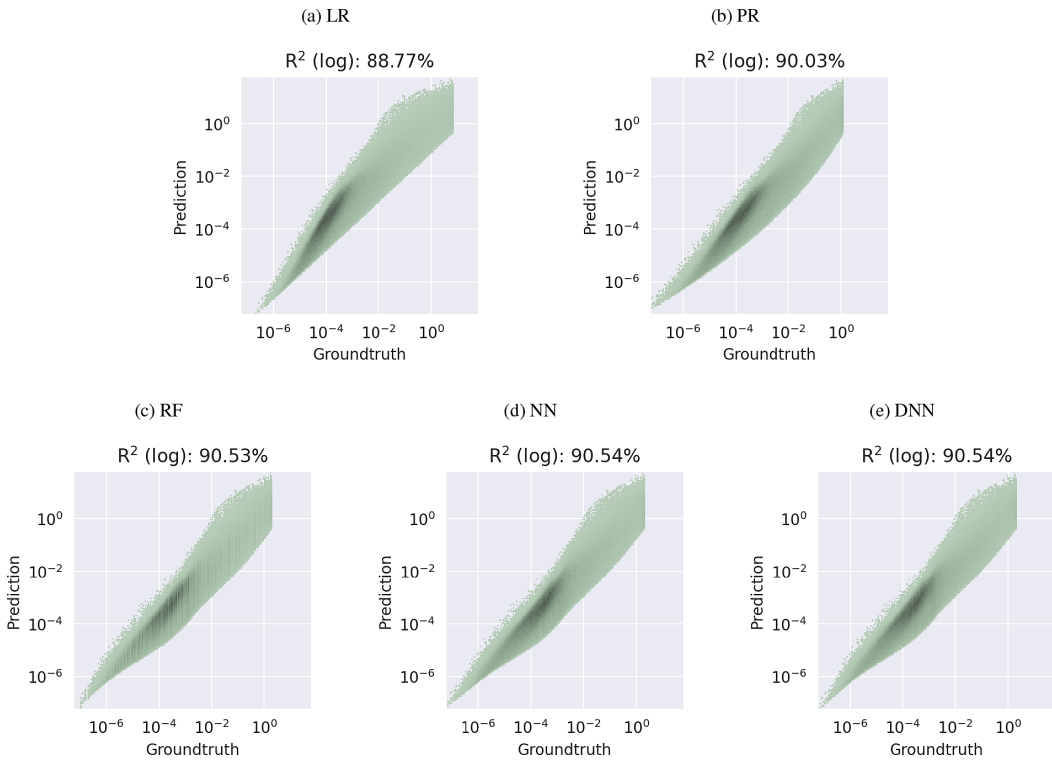
Figure 3. This figure displays five regression plots that compare the predicted and reference simulation autoconversion rates ($g\, m^{-3}\, h^{-1}$) for various machine learning models—(a) linear regression, (b) second-order polynomial regression, (c) random forest, (d) shallow neural network, and (e) deep neural network —on the ICON-LEM Germany testing scenario (scenario 1). The x-axis shows the reference simulation values, while the y-axis shows the predicted autoconversion rates.

input and output. In our particular scenario, where the value of *a* is exceptionally small, approaching zero, we choose to exclude the coefficient *a*. Consequently, with the normalized logarithmic cloud effective radius $\hat{r}_e^{\log}$, we can express the normalized logarithmic autoconversion rates in the form of linear regression as:

$$a_u^{\log} = 0.944\,\hat{r}_e^{\log}.$$

Similar to linear regression, following the second-order polynomial regression formulation $y = a_0 + a_1 x + a_2 x^2$, we obtain the coefficients a_0, a_1, and a_2 after training the model. Consequently, we can also express the normalized logarithmic autoconversion rates using second-order polynomial regression in the following way:

$$a_u^{\log} = 0.085 + 1.009\,\hat{r}_e^{\log} - 0.085\left(\hat{r}_e^{\log}\right)^2.$$

Finally, we can convert the normalized logarithmic autoconversion rates back to their original form by applying the following equation:

$$a_u = 10^{1.396 a_u^{\log} - 9.814}.$$

The units represented for the above equations use $kg\, m^{-3}\, s^{-1}$ for the autoconversion rates and m for the cloud effective radius. In summary, we have outlined simple linear and polynomial regression models

Table 4. *Evaluation of the autoconversion prediction results on simulation model—ICON-LEM Germany and ICON-NWP Holuhraun—using NN model. Testing set 1 is based on ICON-LEM Germany dataset at 1 km resolution. Testing sets 2 and 3 use Cloud-top ICON-LEM Germany at 1 km resolution and Coarser Cloud-top ICON-LEM Germany, respectively. Testing set 4 uses Cloud-top ICON-NWP Holuhraun at 2.5 km resolution*

Testing Set	Time	R2	MAPE	RMSPE	SSIM	PSNR (dB)
1	2 May 2013 1:20 pm	90.54%	9.15%	11.35%	90.11%	26.31
2	2 May 2013 11:40 am	89.10%	10.53%	13.38%	89.20%	25.76
2	2 May 2013 1:20 pm	89.87%	10.88%	13.99%	89.97%	25.89
3 (2.5 km)	2 May 2013 1:20 pm	89.87%	11.46%	14.80%	90.76%	25.45
3 (10 km)	2 May 2013 1:20 pm	89.63%	14.74%	19.63%	91.43%	24.20
4	1 September 2014 7 am	78.89%	11.73%	14.45%	90.00%	23.93
4	1 September 2014 1 pm	84.35%	8.64%	11.65%	91.51%	25.80
4	2 September 2014 11 am	77.35%	11.85%	14.18%	90.61%	23.80
4	3 September 2014 10 am	80.31%	12.27%	14.71%	90.66%	23.68
4	4 September 2014 2 pm	83.96%	9.30%	11.21%	90.99%	24.79
4	5 September 2014 1 pm	79.28%	12.10%	14.47%	90.60%	24.02
4	6 September 2014 5 pm	79.78%	8.99%	10.80%	90.68%	25.00
4	7 September 2014 8 am	79.97%	11.31%	13.74%	89.88%	24.11

that can predict the autoconversion rates using cloud effective radius as the input variable. By utilizing logarithmic transformation and normalization techniques, we can obtain more accurate results.

Cloud-top ICON-LEM Germany: The autoconversion rate prediction metrics using the NN model for the cloud-top ICON-LEM Germany testing scenario are reported in Table 4 under testing set 2. Both R2 and SSIM exhibit values close to 90%, suggesting that one can potentially estimate the auto-conversion rates using model-simulated satellite data accurately, by directly reusing the NN model without further adjustments, such as fine-tuning. This is significant because it alleviates the need to collect additional data and/or run additional time-consuming training processes.

Coarser Horizontal Resolution of Cloud-top ICON-LEM Germany: The autoconversion rate prediction metrics using the NN model for the cloud-top Germany with coarser resolution testing scenarios, 2.5 km and 10 km, are also reported in Table 4. The values of the prediction metrics for the coarser resolution scenarios closely resemble those of the higher resolution (1.0 km). Specifically, for the 2.5 km resolution, MAPE, RMSPE, and PSNR are slightly worse, but SSIM is slightly higher. In the case of the 10 km resolution, MAPE and RMSPE show lower performance compared to the other resolutions; however, the overall results remain quite good, with SSIM performing better. These findings suggest that our machine learning model is capable of working with different data resolutions. This is also significant because atmospheric simulation or satellite data can have very different spatial resolutions since the resolution depends on the simulation models used to generate the data or the satellite instruments used to acquire the data, therefore it is important that the model delivers accurate predictions in the presence of data with different spatial resolutions.

Cloud-top ICON-NWP Holuhraun: We also infer from Table 4 that our NN model predictive performance on the cloud-top ICON-NWP Holuhraun scenario is slightly inferior to that on the cloud-top ICON-LEM Germany scenarios, but it is still fairly decent. This performance decrease is comprehensible given that we trained our model exclusively using the ICON-LEM Germany data and not the ICON-NWP Holuhraun data. However, we conclude we can still use the model trained on the original ICON-LEM Germany data on the ICON-NWP Holuhraun data with reasonably good outcomes, without additional model retraining, demonstrating the model is reasonably robust in this out-of-distribution generalization scenario.

Finally, we visually display the predicted and reference simulation autoconversion rates on a 2D map (for each latitude and longitude) for the ICON-LEM Germany and ICON-NWP Holuhraun testing scenarios in Figures 4 and 5, respectively.

Figure 4a, 4b, and 4c demonstrate that the predicted autoconversion 2D maps closely resemble the reference simulation maps in the various ICON-LEM Germany testing scenarios (1 km, cloud-top 1 km, and cloud-top 2.5 km). Furthermore, Figures 5a, 5b, and 5c also demonstrate that the predicted autoconversion maps exhibit similar characteristics to the reference simulation maps in the ICON-NWP Holuhraun testing scenarios.

Upon observing the difference figures of each scenario depicted on the right-hand side of Figures 4 and 5, a discernible pattern emerges, where our model exhibits a slight tendency to over-predict, as indicated by the predominance of the blue color surpassing the red. Nonetheless, it becomes evident from these visualizations that the variance between the reference simulation and the model's predictions
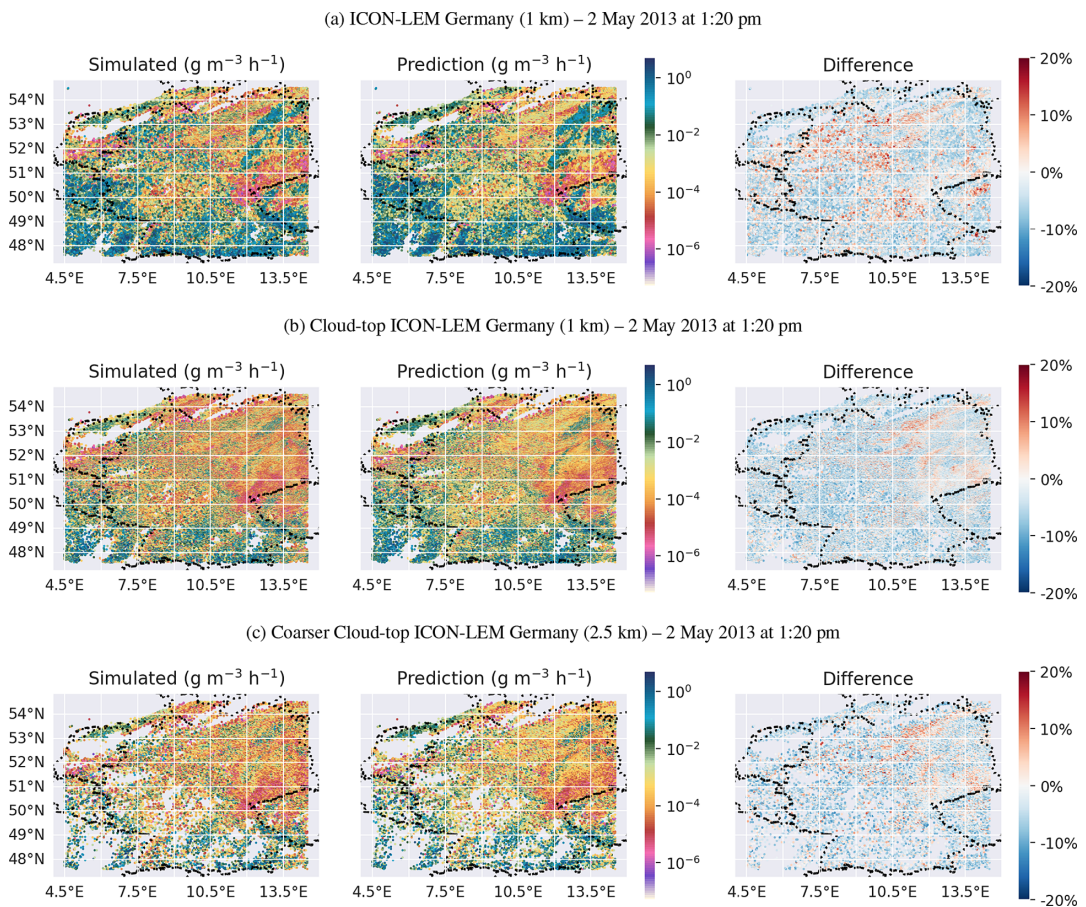
Figure 4. *Visualization of the autoconversion prediction results of ICON-LEM Germany. The left side of the image depicts the reference simulation (simulated), while the middle side shows the prediction results obtained from the shallow NN model. The right side displays the difference between the reference simulation and the prediction results. The top image (a) compares the reference simulation and predictions from ICON-LEM Germany at a resolution of 1 km, depicting each level at different heights in a single 2D map canvas, while the second image (b) focuses on cloud-top information only at a resolution of 1 km. The third figure (c) illustrates the comparison of reference simulation and predictions from ICON-LEM Germany at a coarser resolution of 2.5 km with a focus on cloud-top information only.*
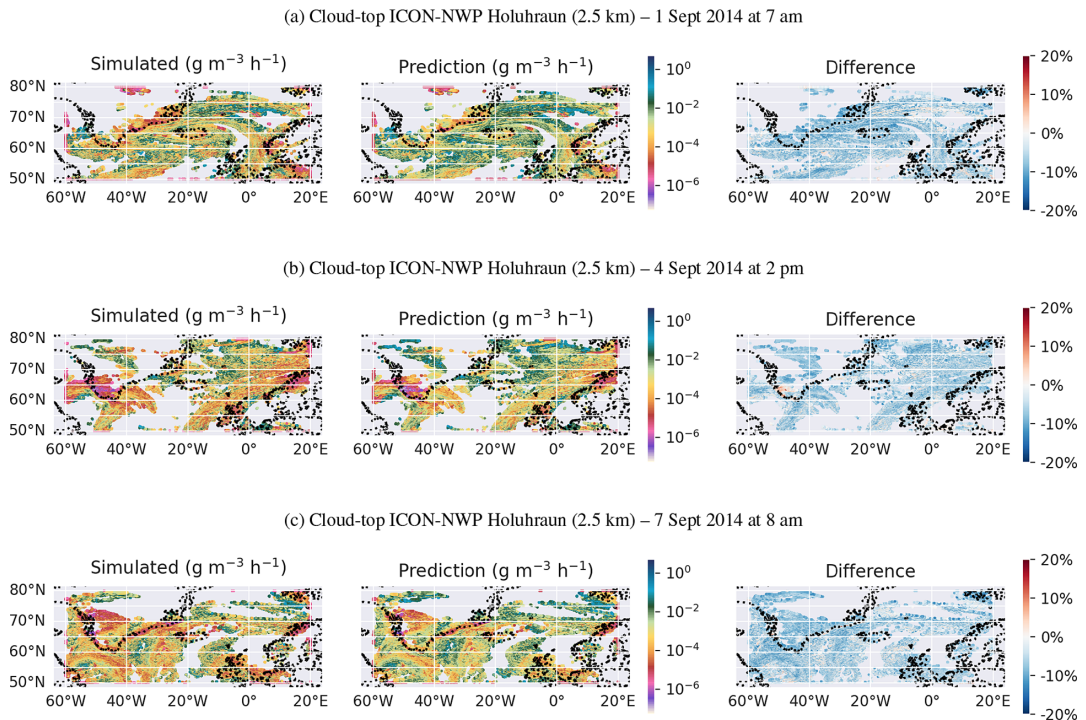
Figure 5. *Visualization of the autoconversion prediction results of ICON-NWP Holuhraun. The left side of the image depicts the reference simulation (simulated), while the middle side shows the prediction results obtained from the shallow NN model. The right side displays the difference between the reference simulation and the prediction results. These three figures provide a comparison between reference simulation and predictions of the ICON-NWP Holuhraun data with a horizontal resolution of 2.5 km, focusing on cloud-top information only, and featuring different time periods.*

remains relatively moderate, with the majority showcasing deviations of less than 10%. Though certain regions exhibit disparities exceeding 10%, the overall performance showcases a reasonable closeness to the actual values. It also reveals that the ICON-NWP Holuhraun testing results exhibit a slightly higher level of over-prediction compared to the ICON-LEM Germany ones. However, as mentioned before, it is comprehensible since we exclusively trained our model using the ICON-LEM Germany data and not the ICON-NWP Holuhraun data.

Overall, these results suggest that the NN model can deliver predictions that are close to the reference simulation data, retaining key spatial features as shown in Figures 4 and 5, in a variety of scenarios, including atmospheric simulation models and satellite-like data.

3.1.1. Comparison with another parameterization

We include a comparison of the autoconversion rates parameterization in our approach with the reference autoconversion rates from Seifert and Beheng [2006], as well as the autoconversion rates parameterization by Freud and Rosenfeld [2012]. We specifically compare with Freud and Rosenfeld [2012] because they also utilize cloud effective radius as an input to derive the parameterization of the mean collection kernel, which is then used to obtain the autoconversion rates. The comparison is conducted for both cloud-top Germany (2 May 2013 at 1:20 pm) and cloud-top Holuhraun (7 September 2014 at 8 am) and is illustrated in Figure 7. From the figure, it is evident that our approach closely aligns with Seifert and Beheng [2006] since we utilize their parameterization as our reference for autoconversion rates. Conversely, the study by Freud and Rosenfeld [2012] results in higher autoconversion rates.
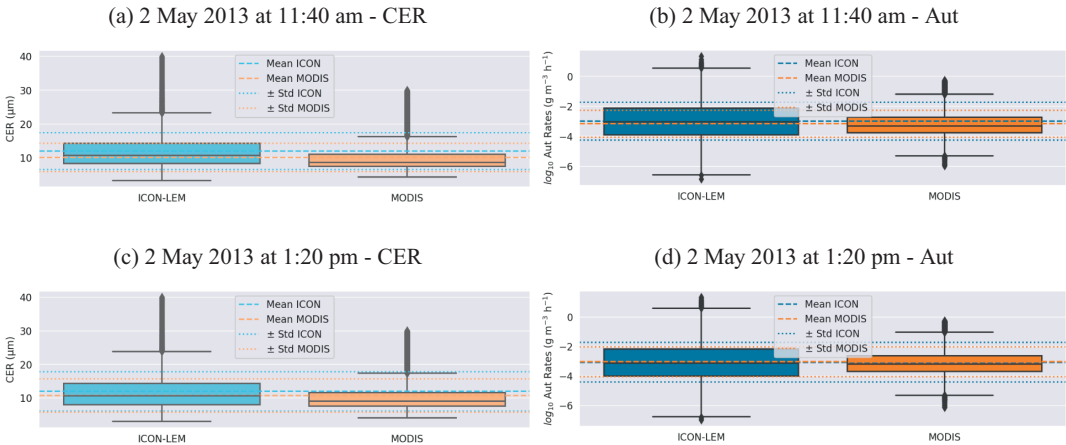
(a) 2 May 2013 at 11:40 am - CER

(b) 2 May 2013 at 11:40 am - Aut

(c) 2 May 2013 at 1:20 pm - CER

(d) 2 May 2013 at 1:20 pm - Aut

Figure 6. *Mean, standard deviation, median, 25th, and 75th percentiles of the cloud-top variables of both ICON and MODIS over Germany. Cloud effective radius (CER) is displayed on the left-hand side in subfigures (a) and (c), while autoconversion rate (Aut) is shown on the right-hand side in subfigures (b) and (d).*
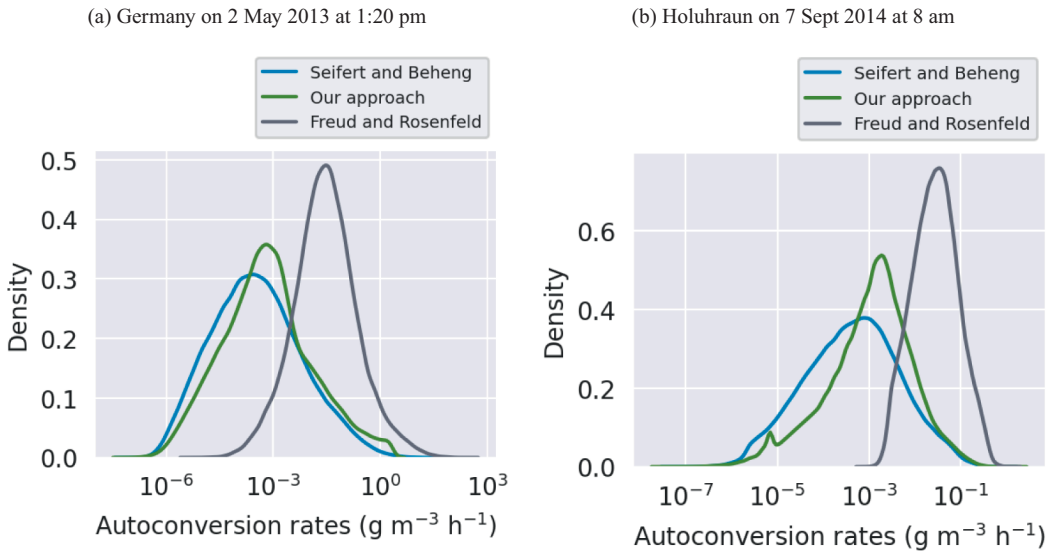

(a) Germany on 2 May 2013 at 1:20 pm

(b) Holuhraun on 7 Sept 2014 at 8 am

Figure 7. *Comparison of the probability density function of autoconversion rates with different parameterizations.*

3.2. *Autoconversion on satellite observations (MODIS)*

We also recall that in this scenario we have used the best machine learning model—identified in the previous section—to predict the autoconversion rates from real satellite MODIS data. In particular, we have chosen to use the NN model as argued earlier.

We compare the autoconversion rate predictions from MODIS satellite data to the cloud-top ICON simulation results relating to the Germany and Holuhraun scenarios.

Figure 6 demonstrates that the MODIS autoconversion rate predictions show statistical agreement with the cloud-top ICON-LEM autoconversion rates over Germany. Specifically, the mean, standard deviation, median, 25th and 75th percentiles of the autoconversion rates of the cloud-top ICON-LEM are relatively close to those of the MODIS over Germany. Minor discrepancies in the percentiles arise due to the cloud
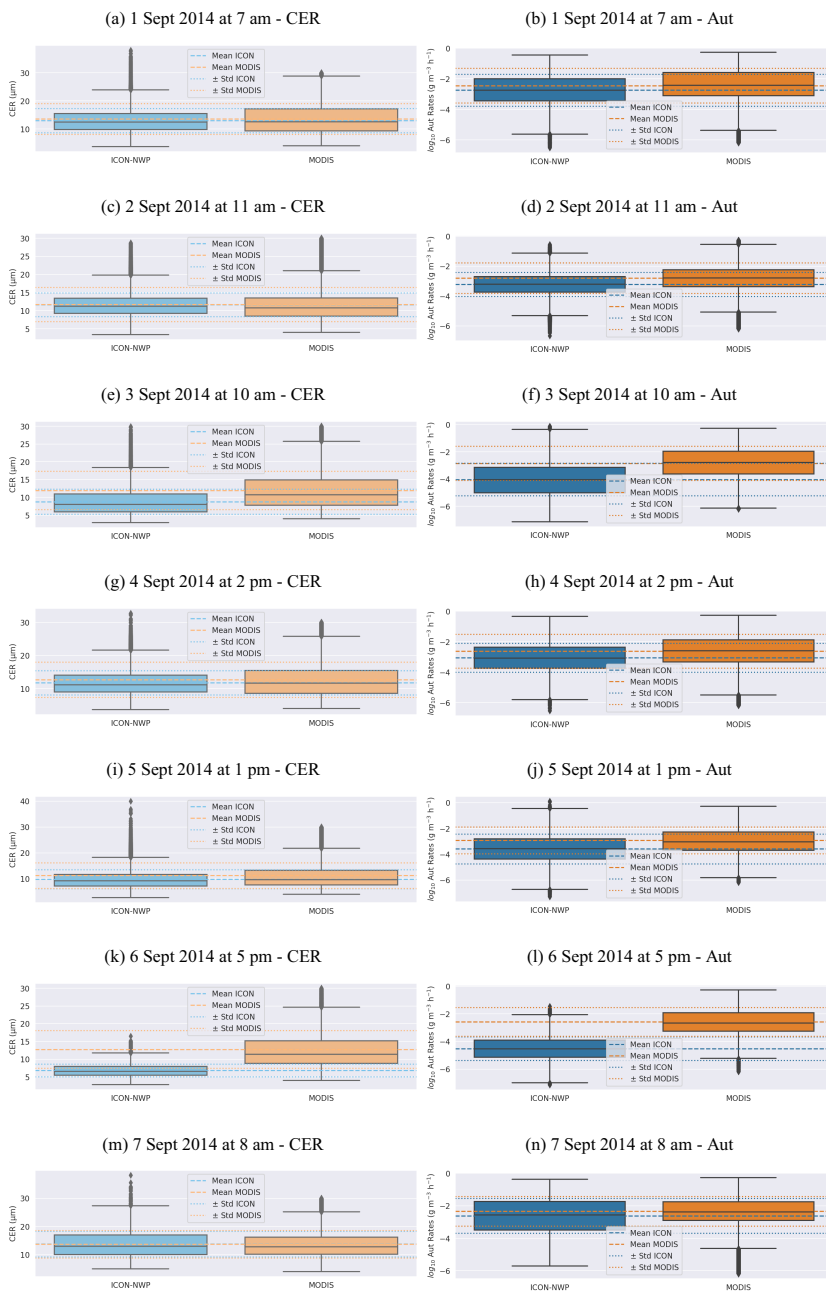
Figure 8. *Mean, standard deviation, median, 25th, and 75th percentiles of the cloud-top variables of both ICON and MODIS over Holuhraun. Cloud effective radius (CER) is displayed on the left-hand side in subfigures (a), (c), (e), (g), (i), (k), and (m), while autoconversion rate (Aut) is shown on the right-hand side in subfigures (b), (d), (f), (h), (j), (l), and (n).*

effective radius of the ICON-LEM model, which tends to yield slightly higher percentiles compared to MODIS. However, the overall agreement in the prediction of autoconversion rates between ICON-LEM and MODIS over Germany remains notably close.

In the case of Holuhraun, as illustrated in Figure 8, most of the autoconversion rates exhibit good agreement between ICON-NWP Holuhraun and MODIS. However, some discrepancies are observed, varying in degree across the rates. Notably, the discrepancies in autoconversion rates (on the right side of the images) between ICON-NWP Holuhraun and MODIS can be attributed to varying discrepancies in
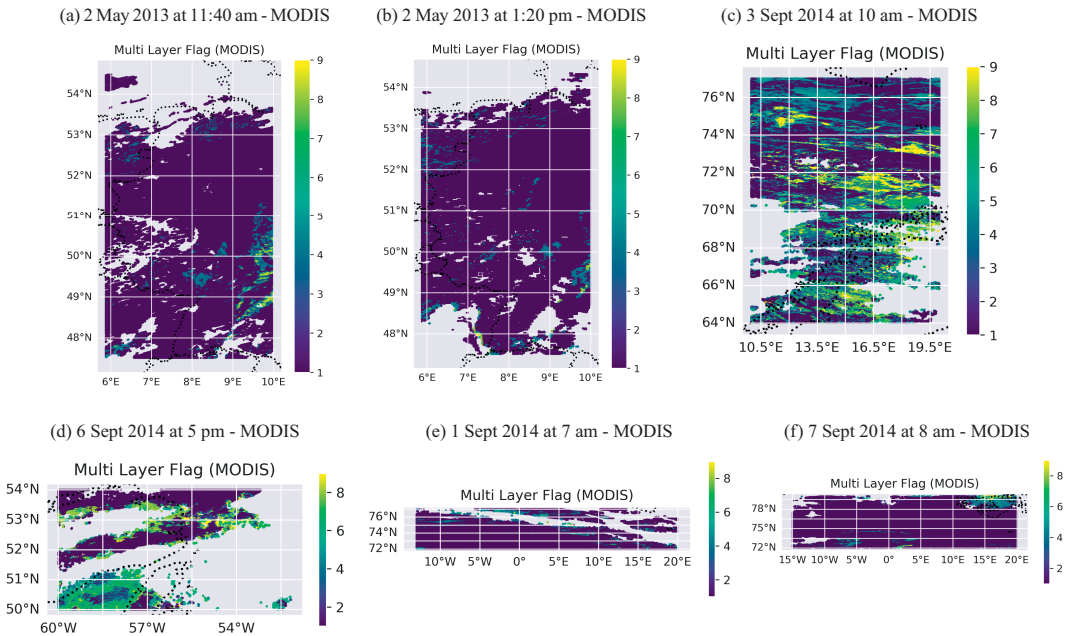
Figure 9. *Visual representation of cloud multi-layer flag of MODIS over Germany and Holuhraun.*

cloud effective radius (on the left side of the images) between ICON-NWP Holuhraun and MODIS. The extent of this variation relies on the magnitude of the difference between simulations (ICON) and satellite data (MODIS). Specifically, a higher discrepancy in cloud effective radius between ICON and MODIS corresponds to a higher discrepancy in autoconversion rates. However, this outcome aligns with our expectations. Since autoconversion is defined based on input values of cloud effective radius, the output will inherently mirror the characteristics of the input. This phenomenon is also observable in the probability density function of the cloud effective radius and the corresponding autoconversion rates at the cloud-top levels of ICON and MODIS over Germany and Holuhraun, as depicted in Figures 10 and 11. These plots showcase the output reflecting input characteristics, although not identical, demonstrating the inherent influence of input on the output.

As previously depicted in Figures 6 and 8, there are larger discrepancies in the mean and standard deviation between ICON and MODIS in Holuhraun compared to those observed in Germany, where the mean and standard deviation difference is relatively lower. A closer examination of specific dates in Holuhraun, particularly on 3 September 2014 and 6 September 2014, reveals substantially larger discrepancies compared to other days. A discernible trend emerges where these discrepancies predominantly occur in conditions where multi-layer clouds are detected, as depicted in Figure 9. From this figure, it can be seen that on days such as 2 May 2013, 1 September 2014, and 7 September 2014, where multi-layer clouds are less prevalent compared to 3 and 6 September 2014, there is better agreement in terms of statistical properties between ICON and MODIS.

In terms of bias, we acknowledge that every simulation and satellite retrieval inherently possesses its own biases compared to actual occurrences. Similarly, our findings also have limitations; the bias in predicting autoconversion rates directly from the satellite, heavily relies on the accuracy of cloud effective radius retrieval from the satellite itself. Particularly, larger discrepancies between cloud effective radius in simulations and satellites directly impact the results of autoconversion rates. This limitation, as demonstrated earlier, is notably prevalent in conditions when multi-layer clouds are dominantly detected. However, despite this constraint, our study suggests a reasonable estimation of autoconversion rates from satellite-derived cloud effective radius or liquid cloud particle size. These findings serve as a starting point for broader discussions and potential avenues for future research.
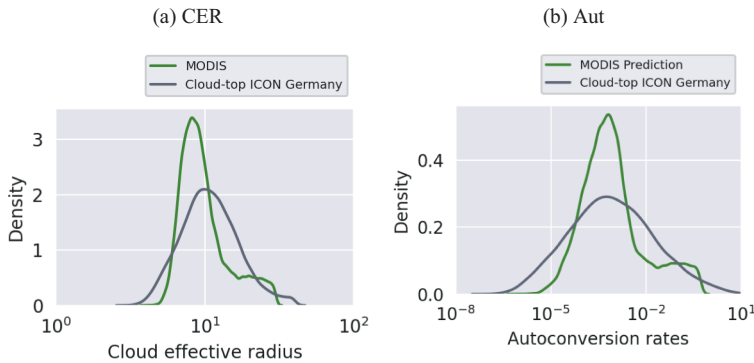
Figure 10. The probability density function of Cloud-top ICON and MODIS variables over Germany on 2 May 2013 at 1:20 pm: cloud effective radius (CER in μm) and autoconversion rates (Aut in $g\,m^{-3}\,h^{-1}$).
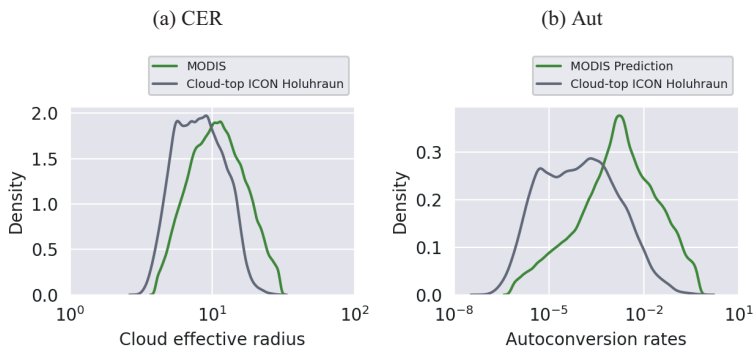

Figure 11. The probability density function of Cloud-top ICON and MODIS variables over Holuhraun on 3 September 2014 at 10 am: cloud effective radius (CER in μm) and autoconversion rates (Aut in $g\,m^{-3}\,h^{-1}$).

4. Concluding remarks

In this study, we provide a computationally effective solution to unravel the key process of precipitation formation for liquid clouds, the autoconversion process, using a machine learning approach. This process is critical to advance our understanding of the response of clouds to anthropogenic aerosols (Mülmenstädt et al. [2020]).

Specifically, we demonstrate that one can reasonably predict the autoconversion rates from cloud effective radius, both from atmospheric simulation data or satellite data, using various machine learning techniques, from the most basic ones such as linear regression to more complex ones such as neural networks. Therefore, our findings confirm that cloud effective radius plays a significant role in determining autoconversion rates in warm phase clouds, confirming other existing results derived from a purely theoretical consideration and in-situ observations (e.g., Freud and Rosenfeld [2012]).

Our work also points to a few directions for future research. First, one could potentially aim to develop other machine learning models that predict autoconversion rates based on additional features beyond the cloud effective radius, such as the cloud optical thickness per layer. Our suggestion is not limited solely to cloud optical thickness per layer; it can also potentially apply to features such as cloud droplet number concentrations per layer and liquid water content per layer. However, it is important to note that currently, these variables per layer are not available in satellite data. Therefore, in order to pursue this approach, it would be necessary to first predict, for instance, the cloud optical thickness per layer and subsequently incorporate it as an additional feature in the prediction of autoconversion rates. Second, one could also

potentially augment the machine learning models with uncertainty quantification capability, in order to deliver not only a point prediction of the variable of interest but also a confidence interval.

Author contribution. Conceptualization: all authors; Supervision: J.Q., M.R.; Methodology: all authors; Visualization: M.N.; Writing—original draft: M.N.; Writing—review and editing: all authors. All authors approved the final submitted draft.

References

Albrecht BA. (1989) Aerosols, cloud microphysics, and fractional cloudiness. *Science*, *245*(4923):1227–1230, ISSN 0036-8075. doi: 10.1126/science.245.4923.1227. URL https://science.sciencemag.org/content/245/4923/1227.

Alfonso L and Zamora JM. (2021) A two-moment machine learning parameterization of the autoconversion process. *Atmospheric Research*, *249*:105269, ISSN 0169-8095. doi: https://doi.org/10.1016/j.atmosres.2020.105269. URL https://www.sciencedirect.com/science/article/pii/S0169809520312060.

Bellouin N, **Quaas J**, **Gryspeerdt E**, **Kinne S**, **Stier P**, **Watson-Parris D**, **Boucher O**, **Carslaw KS**, **Christensen M, and Daniau A-L**, et al. (2020). Bounding global aerosol radiative forcing of climate change. *Reviews of geophysics.*, *58*(1), ISSN 8755-1209. doi: 10.1029/2019RG000660.

Breiman L. Random forests.(2001) *Machine Learning*, *45*(1):5–32,. ISSN 0885-6125. doi: 10.1023/A:1010933404324. URL https://doi.org/10.1023/A

Chiu JC, **Yang CK**, **van Leeuwen PJ**, **Feingold G**, **Wood R**, **Blanchard Y**, **Mei F, and Wang J**. (2021) Observational constraints on warm cloud microphysical processes using machine learning and optimization techniques. *Geophysical Research Letters*, *48*(2), 1. doi: 10.1029/2020GL091236.

Costa-Surós M, **Sourdeval O**, **Acquistapace C**, **Baars H**, **Henken CC**, **Genz C**, **Hesemann J**, **Jimenez C**, **König M, and Kretzschmar J**, et al. (2020) Detection and attribution of aerosol–cloud interactions in large-domain large-eddy simulations with the icosahedral non-hydrostatic model. *Atmospheric Chemistry and Physics*, *20*(9): 5657–5678, doi: 10.5194/acp-20-5657-2020. URL https://acp.copernicus.org/articles/20/5657/2020/.

Czernia D and Szyk B. (2021) Air density calculator. Available at https://www.omnicalculator.com/physics/air-density (accessed 18 September 2021).

Dipankar A, **Stevens B**, **Heinze R**, **Moseley C**, **Zängl G**, **Giorgetta M, and Brdar S**. (2015) Large eddy simulation using the general circulation model icon. *Journal of Advances in Modeling Earth Systems*, *7* (3):963–986, doi: https://doi.org/10.1002/2015MS000431. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015MS000431.

Freud E and Rosenfeld D. (2012) Linear relation between convective cloud drop number concentration and depth for rain initiation. *Journal of Geophysical Research: Atmospheres*, *117* (D2),. doi: https://doi.org/10.1029/2011JD016457. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JD016457.

Gryspeerdt E, **Goren T**, **Sourdeval O**, **Quaas J**, **Mülmenstädt J**, **Dipu S**, **Unglaub C**, **Gettelman A, and Christensen M**. (2019) Constraining the aerosol influence on cloud liquid water path. *Atmospheric Chemistry and Physics*, *19*(8): 5331–5347, doi: 10.5194/acp-19-5331-2019. URL https://acp.copernicus.org/articles/19/5331/2019/.

Haghighatnasab M, **Kretzschmar J**, **Block K, and Quaas J**. (2022) Impact of holuhraun volcano aerosols on clouds in cloud-system-resolving simulations. *Atmospheric Chemistry and Physics*, *22*(13): 8457–8472,. doi: 10.5194/acp-22-8457-2022. URL https://acp.copernicus.org/articles/22/8457/2022/.

Heinze R, **Dipankar A**, **Henken CC**, **Moseley C**, **Sourdeval O**, **Trömel S**, **Xie X**, **Adamidis P**, **Ament F, and Baars H**, et al. (2017) Large-eddy simulations over Germany using ICON: a comprehensive evaluation. *Q. J. R. Meteorol. Soc.*, *143*(702): 69–100, doi: 10.1002/qj.2947.

IPCC. (2021) Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen,. *Cambridge Univ. Press*, (In Press):3949, URL https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Full_Report.pdf.

Marshall JS and Palmer WMcK. (1948)The distribution of raindrops with size. *Journal of Atmospheric Sciences*, 5(4):165 – 166, doi: 10.1175/1520-0469(1948)005<0165:TDORWS>2.0.CO;2. URL https://journals.ametsoc.org/view/journals/atsc/5/4/1520-0469_1948_005_0165_tdorws_2_0_co_2.xml.

McCulloch WS and Pitts W. (1943) A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5* (4):115–133.

Mülmenstädt J, **Nam C**, **Salzmann M**, **Kretzschmar J**, **L'Ecuyer TS**, **Lohmann U**, **Ma P**, **Myhre G**, **Neubauer D**, **Stier P**, et al. (2020) Reducing the aerosol forcing uncertainty using observational constraints on warm rain processes. *Science Advances*, *6* (22):eaaz6433, doi: 10.1126/sciadv.aaz6433. URL https://www.science.org/doi/abs/10.1126/sciadv.aaz6433.

O'Malley T, **Bursztein E**, **Long J**, **Chollet F**, **Jin H**, **Invernizzi L**, et al. (2019) Keras Tuner. https://github.com/keras-team/keras-tuner.

Pinsky M, **Khain AP, and Shapiro M**. (2001) Collision efficiency of drops in a wide range of reynolds numbers: Effects of pressure on spectrum evolution. *J. Atmos. Sci.*, *58*(7):742–755, ISSN 00224928. doi: 10.1175/1520-0469(2001)058<0742:ceodia>2.0.co;2.

Platnick S. (2000) Vertical photon transport in cloud remote sensing problems. *Journal of Geophysical Research: Atmospheres*, *105* (D18):22919–22935. doi: https://doi.org/10.1029/2000JD900333. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000JD900333.

Platnick S, **Meyer KG**, **King MD**, **Wind G**, **Amarasinghe N**, **Marchant B**, **Arnold GT**, **Zhang Z**, **Hubanks PA**, **Ridgway B, and Riedi J**. (2018) MODIS Cloud Optical Properties: User Guide for the Collection 6/6.1 Level-2 MOD06/MYD06 Product and Associated Level-3 Datasets. URL https://modis-atmos.gsfc.nasa.gov/sites/default/files/ModAtmo/MODISCloudOpticalPropertyUserGuideFinal_v1.1_1.pdf.

Platnick S, **Meyer KG**, **King MD**, **Wind G**, **Amarasinghe N**, **Marchant B**, **Arnold GT**, **Zhang Z**, **Hubanks PA**, **Holz RE**, **Yang P**, **Ridgway WL, and Riedi J**. (2017) The MODIS Cloud Optical and Microphysical Products: Collection 6 Updates and Examples from Terra and Aqua. *IEEE Trans. Geosci. Remote Sens.*, *55*(1): 502–525, doi: 10.1109/TGRS.2016.2610522.

Reuther A, **Michaleas P**, **Jones M**, **Gadepally V**, **Samsi S, and Kepner J**. (2019) Survey and benchmarking of machine learning accelerators. In *2019 IEEE High Performance Extreme Computing Conference (HPEC)*, pages 1–9,. doi: 10.1109/HPEC.2019.8916327.

Rolnick D, **Donti PL**, **Kaack LH**, **Kochanski K**, **Lacoste A**, **Sankaran K**, **Ross AS**, **Milojevic-Dupont N**, **Jaques N, and Waldman-Brown A**, et al. (2019). Tackling climate change with machine learning. *Association for Computing Machinery 15* https://doi.org/10.1145/3485128.

Schmidhuber J. (2015) Deep learning in neural networks: An overview. *Neural Networks*, *61*:85 – 117,. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2014.09.003. URL http://www.sciencedirect.com/science/article/pii/S0893608014002135.

Seifert A and Beheng KD. (2006) A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 1: Model description. *Meteorol. Atmos. Phys.*, *92*(1-2):45–66, ISSN 01777971. doi: 10.1007/s00703-005-0112-4.

Seifert A and Rasp S. (2020) Potential and Limitations of Machine Learning for Modeling Warm-Rain Cloud Microphysical Processes. *J. Adv. Model. Earth Syst.*, *12*(12), ISSN 19422466. doi: 10.1029/2020MS002301.

Stephan Kolzenburg, DG, **Thordarson T**, **Höskuldsson A, and Dingwell DB**. (2017) The rheological evolution of the 2014/2015 eruption at holuhraun, central iceland. *Bulletin of Volcanology*, *79*(6):1–16.

Stevens B, **Acquistapace C**, **Hansen A**, **Heinze R**, **Klinger C**, **Klocke D**, **Rybka H**, **Schubotz W**, **Windmiller J, and Adamidis P**, et al. (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, doi: 10.2151/jmsj.2020-021. URL https://doi.org/10.2151/jmsj.2020-021.

Twomey. S. Pollution and the planetary albedo. *Atmospheric Environment (1967)*, *8*(12): 1251–1256, 1974. ISSN 0004-6981. doi: https://doi.org/10.1016/0004-6981(74)90004-3. URL https://www.sciencedirect.com/science/article/pii/0004698174900043.

Zängl G, **Reinert D**, **Rípodas P, and Baldauf M**. (2015)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, ISSN 1477870X. doi: 10.1002/qj.2378.

Zhang Z and Platnick S. (2011) An assessment of differences between cloud effective particle radius retrievals for marine water clouds from three MODIS spectral bands. *Journal of Geophysical Research (Atmospheres)*, *116* (D20):D20215, October. doi: 10.1029/2011JD016216.