An ML-based P3-like multimodal two-moment ice microphysics in the ICON model

Axel Seifert[1] and Christoph Siewert[1]

[1]Deutscher Wetterdienst

April 16, 2024

Abstract

Machine learning (ML) is used to build a bulk microphysical parameterization including ice processes. Simulations of the Lagrangian super-particle model McSnow are used as training data. The machine learning performs a coarse-graining of the particle-resolved microphysics to multi-category two-moment bulk equations. Besides mass and number, prognostic particle properties (P3) like melt water, rime mass, and rime volume are predicted by the ML-based bulk model. The ML-based scheme is tested with simulations of increasing complexity. As a box model, the ML-based bulk scheme can reproduce the simulations of McSnow quite accurately. In 3d idealized squall line simulations, the ML-based P3-like scheme provides a more realistic extended stratiform region when compared to the standard two-moment bulk scheme in ICON. In a realistic case study, the ML-based scheme runs stably, but can not significantly improve the results. This shows that machine learning can be used to coarse-grain super-particle simulations to a bulk scheme of arbitrary complexity.

An ML-based P3-like multimodal two-moment ice microphysics in the ICON model

Axel Seifert[1], Christoph Siewert[1]

[1]Deutscher Wetterdienst, Offenbach, Germany

Key Points:

- Machine learning is successfully applied to build a complex bulk ice microphysics scheme by coarse-graining output of a Lagrangian particle microphysics model.
- The ML-based P3-like microphysics scheme improves the representation of the stratiform region of an idealized squall line compared to a classic two-moment scheme.
- The ML-based P3-like microphysics scheme runs stable and provides meaningful results in three-dimensional real-case simulations.

Corresponding author: Axel Seifert, `axel.seifert@dwd.de`

Abstract

Machine learning (ML) is used to build a bulk microphysical parameterization including ice processes. Simulations of the Lagrangian super-particle model McSnow are used as training data. The machine learning performs a coarse-graining of the particle-resolved microphysics to multi-category two-moment bulk equations. Besides mass and number, prognostic particle properties (P3) like melt water, rime mass, and rime volume are predicted by the ML-based bulk model. The ML-based scheme is tested with simulations of increasing complexity. As a box model, the ML-based bulk scheme can reproduce the simulations of McSnow quite accurately. In 3d idealized squall line simulations, the ML-based P3-like scheme provides a more realistic extended stratiform region when compared to the standard two-moment bulk scheme in ICON. In a realistic case study, the ML-based scheme runs stably, but can not significantly improve the results. This shows that machine learning can be used to coarse-grain super-particle simulations to a bulk scheme of arbitrary complexity.

Plain Language Summary

Numerical weather prediction and climate models need a description of unresolved cloud microphysical processes. Such microphysical parameterizations are usually formulated as systems of equations for bulk variables that describe the time evolution of clouds and precipitation. In this study, we use machine learning (ML) techniques to build such a parameterization. As input or training data simulations of a very detailed cloud model are used. This detailed model provides information not only on the mass and number of cloud particles but also other properties like the degree of melting or the mass of liquid drops frozen on the ice particles called rime mass. The machine learning approach can successfully construct the necessary statistical relations that are needed for microphysical parameterization. This parameterization is then tested in simulations of increasing complexity. The new ML-based scheme provides physically reasonable solutions and improves the simulation of a line of thunderstorms

1 Introduction

Developing parameterizations for numerical weather prediction (NWP) and climate models can be a tedious and time-consuming task (Jakob, 2010). Speeding up this development cycle is crucial for further progress in understanding and predicting regional climate change and improve NWP models to forecast hazardous and extreme weather events (Bauer, Stevens, & Hazeleger, 2021).

Machine learning methods hold the promise for a more rapid model development cycle, for example, through a semi-automatic workflow from highly-resolved reference simulations to coarse-grained and computationally efficient algorithms. Machine learning algorithms as an integral part of NWP and climate models may allow for better performance optimization and on-the-fly calibration with observations (Bauer, Dueben, et al., 2021). Machine learning methods have recently gained much attention in atmospheric modeling especially for emulators that help to improve the computational performance of the model (Ukkonen et al., 2020; Lagerquist et al., 2021; Chantry et al., 2021; Meyer, Grimmond, et al., 2022; Meyer, Hogan, et al., 2022; Ukkonen, 2022).

Here we take a rather straightforward approach to machine learning in that we use supervised learning with fully connected neural nets applied to individual physical process rates. This approach has the advantage that it is conceptually very similar to classic parameterizations, i.e., the result is an ODE system for the bulk variables (Seifert & Rasp, 2020; Gettelman et al., 2021). It also ensures mass conservation and allows a posteriori analysis of the ML representation of individual physical processes. In addition, it can be applied at different model time steps and even horizontal resolutions without having to re-train the ML model. The main disadvantage of this simple approach to ML is that it is not as computationally efficient as ML methods could be if applied in a more advanced and state-of-the-art framework, e.g., using a UNet++ architecture (Lagerquist et al., 2021) or recurrent neural nets (Ukkonen, 2022). Our ML approach is quite similar to the use of look-up tables for microphysical process rates. The use of look-up tables has a long tradition in cloud microphysical modeling (Walko et al., 1995; Feingold et al., 1998). To avoid the term look-up-table they are sometimes even called bin-emulating schemes in cloud modeling literature (Khain et al., 2015), which should not mask the fact these are still bulk schemes with their intrinsic limitations.

Many of the currently available bin and bulk microphysical schemes have problems in representing the stratiform region of mesoscale convective systems like squall lines (Morrison et al., 2009; Xue et al., 2017). This is often attributed to the fact that they are based on a limited number of particle types like snow and graupel, whereas in nature the transition from snow to graupel by riming, i.e. by accretion of supercooled cloud droplets, is continuous (Mosimann et al., 1994; Seifert et al., 2019). Morrison and Milbrandt (2015, MM15 hereafter) suggested to abandon such particle types completely and instead use prognostic particle properties (P3) especially the rime mass and the rime volume to represent riming. The additional rime volume is important to predict rime density. Instantaneous rime density is a function of temperature and Stokes number (Cober & List, 1993), but for a given particle the rime density depends on its history and therefore requires an additional prognostic variable. The importance of prognostic rime mass for the simulation of deep convection is also discussed in Aligo et al. (2018). In their original P3 scheme, MM15 abandoned the multimodal representation that comes with multiple particle classes, but later they presented a version of their scheme with multiple categories (Milbrandt & Morrison, 2016), and recently also a version of P3 with a triple-moment representation (Milbrandt et al., 2021). In addition, an extended variant of the P3 scheme with prognostic melt water on ice particles has been developed (Cholette et al., 2019, 2020, 2023). Hence, the P3 approach represents the state-of-the-art of bulk microphysical parameterizations for high-resolution NWP and climate models.

In the following, we explore whether we can derive or 'learn' a P3-like scheme from Lagrangian super-particle simulations using standard machine learning methods. The aim is to build a semi-automatic workflow that generates a bulk microphysical scheme based on some a priori choices and simulations of the super-particle model McSnow.

The paper is organized as follows: In section 2 we introduce the basic assumptions of the new ML-based P3-like microphysics scheme. In section 3 the super-particle model McSnow and the simulations that serve as training data are described. The actual structure of the training data and the training process are discussed in section 4. Section 5 presents a comparison of McSnow and the ML-based bulk model. In section 6 the ML-based model is applied to idealized three-dimensional squall line simulations with the ICON model. Section 7 presents a realistic case study with ICON and a comparison of the ML-based bulk models with a classical two-moment bulk microphysics scheme. The paper ends with a Summary and Conclusions.

2 A priori choices for the ML-based P3-like microphysics

To build an ML-based microphysics scheme, we have to make some a priori choices regarding the number of hydrometeor categories and the corresponding prognostic bulk variables. In contrast to MM15, we think that multimodality is ubiquitous in clouds because the various pathways for the formation of precipitation-sized particles lead to the co-existence of various modes or particle types. Hence, our scheme has multiple particle categories, but they have a clear and physically-based definition in terms of their formation mechanism. The cloud ice category comprises primary ice particles (monomers) which have only grown by depositional growth. Unrimed snow are aggregates of these primary crystals. Those two categories have only a two-moment representation with no additional properties. Then we have three categories that carry rime mass and rime volume: rimed ice, rimed snow, and graupel. Whereas rimed ice and rimed snow are simply the rimed monomers and rimed aggregates, the graupel originates from freezing of raindrops.

The latter two categories have a prognostic liquid water mass to explicitly represent melting and wet growth. Carrying unrimed ice (snow) and rimed ice (rimed snow) separately may sound unnecessary, given that we have prognostic rime mass, but due to the patchiness of supercooled liquid water a co-existence of rimed and unrimed particles in the same grid volume is not impossible. The hydrometeor categories and the corresponding prognostic variables of the ML-based P3-like scheme are summarized in Table 2. The scheme has overall 23 prognostic variables: 18 for the ice phase, 4 for the liquid phase, and one additional tracer for tracking activated ice nuclei (Köhler & Seifert, 2015). Note that the bulk classification in McSnow is different from the classes of the ML-based P3-like scheme in that the McSnow classification would allow a conversion from snow to graupel. In fact, this process is contained in the training data, but for the ML-based P3-like bulk scheme described in the current study, we decided not to allow snow-to-graupel conversion. Note that the particle classification in McSnow is only a diagnostic to analyze the simulations and generate training data for a bulk model. The bulk classification does not affect the microphysical processes in McSnow, which is by construction continuous and class-free.

With 23 prognostic variables and a high level of complexity, this ML-based scheme is not primarily aimed at operational NWP, where computational efficiency is of the essence

Table 1. Prognostic variables of the super-particle model McSnow and corresponding bulk variables of the ML-based P3-like scheme. Here ρ_i is the material density of ice and χ is the super-particle multiplicity divided by air volume. The hydrometeor categories are defined in Table 2.

Prognostic variables of McSnow super-particles		
variable	symbol	note
ice mass	m_i	increases by depositional growth
rime mass	m_r	increases by riming
rime volume	V_r	changes obey instantaneous rime density
liquid mass	m_ℓ	increases by melting or collection of liquid drops
frozen mass	m_f	increases by freezing of liquid mass
monomer number	N	number of ice monomers
multiplicity	χ	number of real particles per super-particle
Prognostic variables of ML-based P3-like bulk scheme for a hydrometeor category k		
mass density	q_k	sum of $(m_i + m_r + m_f + m_\ell)\chi$
number density	n_k	sum of χ
rime mass	ψ_k	sum of $(m_r + m_f)\chi$
rime volume	ϕ_k	sum of $(V_r + m_f/\rho_i)\chi$
liquid mass	ℓ_k	sum of $m_\ell\chi$

but in cloud modeling, regional climate research and other applications that may care about a good representation of cloud microphysics. The large number of prognostic variables make this scheme rather complicated, but quite suitable as a test case and proof-of-concept for the ML approach to parameterization development.

3 Super-particle simulations

The Lagrangian super-particle model McSnow (Brdar & Seifert, 2018) makes use of the Monte-Carlo algorithm of Shima et al. (2009) to simulate the collision and aggregation processes of hydrometeors. The super-particle approach allows for a direct representation of the evolution of the properties of individual hydrometeors. To do so, McSnow carries multiple variables to describe each hydrometeor. These are the hydrometeor state variables ice mass, rime mass, rime volume, number of monomers, frozen mass, and liquid mass (see Table 1). The ice mass increases due to depositional growth and determines the maximum dimension of the particle with the help of an empirical m-D relationship. Hence, in this configuration of McSnow, we do not employ the habit pre-

Table 2. Overview of the ML-based P3-like two-moment bulk microphysics scheme. The degree of riming ξ is here defined as $\xi = (m_r + m_f)/(m_i + m_r + m_f + m_\ell)$ including the frozen mass m_f.

class	variables	McSnow classification
unrimed ice	q_i, n_i	$N = 1$ and $\xi = 0$
unrimed snow	q_s, n_s	$N > 1$ and $\xi = 0$
rimed ice	q_{ri}, n_{ri}, ψ_{ri}, ϕ_{ri}	$N = 1$ and $0 < \xi \leq 0.95$
rimed snow	q_{rs}, n_{rs}, ψ_{rs}, ϕ_{rs}, ℓ_{rs}	$N > 1$ and $0 < \xi \leq 0.95$
graupel	q_g, n_g, ψ_g, ϕ_g, ℓ_g	$\xi > 0.95$
cloud droplets	q_c, n_c	$m_i + m_r + m_f = 0$ and $r < 40$ μm
raindrops	q_r, n_r	$m_i + m_r + m_f = 0$ and $r \geq 40$ μm

diction of Welss et al. (2023). Rime mass and rime volume increase due to collision with supercooled liquid drops. The instantaneous rime density is parameterized following Cober and List (1993). The number of monomers increases by aggregation, i.e., collection of other ice particles. Finally, melting in McSnow is based on Rasmussen et al. (1984a, 1984b) and Rasmussen and Heymsfield (1987b, 1987a). Melting increases the liquid mass m_ℓ, which is a prognostic variable for each individual super-particle. Freezing of liquid drops uses a probabilistic interpretation of the parameterization of Barklie and Gokhale (1959). As secondary ice production, only Hallet-Mossop rime splintering is currently considered in McSnow (Hallett & Mossop, 1974; Field et al., 2017).

We assume that all hydrometeors fall with their terminal fall velocity v_t. The terminal fall velocities of all hydrometeors and also the collision efficiency E_c of all possible mutual collisions are parameterized using the approach of Böhm (1992a, 1992b, 1992c, 1994, 1999, 2004). Using Böhm's theory provides a continuous and physically consistent dependency of the hydrometeor properties, like v_t or E_c, and consequently the collision kernel K on the hydrometeor state variables. Welss et al. (2023) provide a more detailed discussion of Böhm's theory in the framework of McSnow. Special considerations are necessary for the sticking efficiency of unrimed and rimed snow and graupel. Usually, the sticking efficiency of snow is parameterized as a function of temperature whereas the sticking efficiency of graupel is most often assumed to be small and constant. In McSnow a continuous parameterization as a function of temperature and degree of riming is applied, which is specified in Appendix A.

McSnow can be used in a one-dimensional version as in Brdar and Seifert (2018) and Bringi et al. (2020) and in two- and three-dimensional simulation as part of the ICON model (Siewert & Seifert, 2018). As training data for the machine learning approach, we need a broad range of environmental and microphysical parameters. Hence, two- or three-dimensional simulations would be far too expensive. Even the one-dimensional model is inefficient because it needs several hours of simulation time to spin up a quasi-equilibrium state. To overcome these obstacles, we have implemented a simple zero-dimensional box model that approximates the quasi-equilibrium state of the one-dimensional McSnow, but is computationally cheaper. The zero-dimensional McSnow describes a population of hydrometeors initialized as pristine ice of unrimed monomers that fall through a pre-described atmosphere. The sedimentation velocity of the box model is equal to the mass-weighted terminal fall velocity of all hydrometeors in the box. The atmospheric profile is the same as in Brdar and Seifert (2018), their Figure 6. While the box is falling through the atmosphere, the hydrometeors grow by depositional growth and mutual binary collisions. They encounter a layer of supercooled liquid drops and grow by riming. Without the presence of liquid water, ice particles start melting when they reach the 0°C level. As melting in McSnow is formulated by a quasi-equilibrium energy budget, large ice particles can reach the wet growth regime in regions of high liquid water content. In the wet growth regime, a liquid water layer exists on the ice particles even at temperatures below 0 °C.

The simulations described in the previous paragraph mimic the microphysical processes in a stratiform cloud including the stratiform regions of convective systems. In convective updrafts other processes, like freezing of raindrops and riming with raindrops are important or even dominant, which are not well represented in those simulations. To sample the microphysical processes as they occur in convective clouds, the same atmospheric profile is used, but the box model is initialized near the surface with an arbitrarily chosen upward velocity of 5 m/s. This leads to the formation of raindrops in the parcel, which subsequently freeze and start riming. These simulations provide the data for microphysical processes as they happen within updraft cores of convective systems. When the updraft parcel reaches a height of $0.95\,h_{\text{top}}$, the updraft ends, and the parcel enters the regular sedimentation mode described above, where it falls with the mass-weighted sedimentation velocity of the hydrometeors. This is necessary to provide, for example, training data for the melting of graupel.

Table 3. Parameter hypercube of McSnow simulations for the training data. The random sampling is based on uniform random variables $r \in [0.1]$. The parameters for additive sampling are $c_h = 1$ km and $c_r = 5$ μm. Not all heights h_1 and Δh_2 have been used with all domain tops h_{top}.

	Basic McSnow simulations		
variable	range of values	random sampling	unit
ice supersaturation S_i	0.0, 0.2, 0.5	$S_i^* = S_i\,(1+r)$	-
height h_1	500, 1000 1500	$h_1^* = h_1 + C\,r$	m
height Δh_2	500, 1500, 2000	$h_2^* = h_1 + \Delta h_2\,(1+r)$	m
droplet radius r_c	5, 10, 15, 20, 25, 30	$r_c^* = r_c + Cr$	μm
ice number density N_i	10, 20, 40, 80, 160, 320, 640	$N_i^* = N_i(1+r)$	dm^{-3}
ice water content Q_i	0.1, 0.2, 0.4	$Q_i^* = Q_i(1+r)$	g cm^{-3}
cloud water content Q_c	0.1, 0.2, 0.4, 0.8	$Q_c^* = Q_c(1+r)$	g cm^{-3}
domain top h_{top}	5500, 6500, 7500, 8500, 9500	-	m
	Updraft McSnow simulations		
variable	range of values	random sampling	C
ice supersaturation S_i	0.0, 0.2	$S_i^* = S_i\,(1+r)$	-
height h_1	500, 1500, 2500	$h_1^* = h_1 + C\,r$	m
height Δh_2	2000, 4000	$h_2^* = h_1 + \Delta h_2\,(1+r)$	m
droplet radius r_c	5, 10, 15, 20, 25, 30	$r_c^* = r_c + Cr$	μm
ice number density N_i	10, 20, 40, 80, 160, 320, 640	$N_i^* = N_i(1+r)$	dm^{-3}
ice water content Q_i	0.1, 0.2, 0.4	$Q_i^* = Q_i(1+r)$	g cm^{-3}
cloud water content Q_c	0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4	$Q_c^* = Q_c(1+r)$	g cm^{-3}
domain top h_{top}	6000, 7000, 9000	-	m

4 Training data and ML results

To build a bulk microphysics scheme using a standard machine learning workflow, we first choose the prognostic variables of the desired scheme. Here we decided on a two-moment approach with seven particle categories, and rimed ice, rimed snow and graupel have additional prognostic variables for rime mass, rime volume, and liquid mass (see Table 2). As shown by Seifert and Rasp (2020, SR20 hereafter) the ML approach has no advantages for the warm-rain processes that determine the growth of cloud droplets and raindrops. Hence, we use existing parameterizations for the warm-rain processes based on Seifert and Beheng (2001) and Seifert (2008). Ice nucleation is parameterized using semi-empirical approaches and is not modified by machine learning. The ice nucleation active site (INAS) density approach of Ullrich et al. (2017) is used for heterogeneous ice nucleation, whereas homogeneous ice nucleation follows Kärcher et al. (2006).

This leaves 55 process rates that need to be parameterized by the machine learning approach. Those process rates include depositional growth and melting of ice particles, and all collisional processes like aggregation among ice particles and riming of ice by collection of cloud droplets or raindrops. All those processes depend, on one hand, on the physical properties of the hydrometeors (terminal fall velocity, particle size distribution) and change, on the other hand, the bulk properties of the ice categories. Tables 4-6 summarize all those processes, the input variables (predictors or features) of the neural net, and the predicted process rates (output or labels). In addition to the 55 process rates (networks 1-27), the bulk sedimentation velocities are needed to quantify the precipitation fluxes of the particle categories (networks 28-32). To compare with observations, an estimate of the radar reflectivity is needed that depends primarily on the 2nd mass moment. Hence, this also has to be estimated by a neural net (networks 33-37). Finally, for a consistent coupling with radiation, we need effective radii for each category (networks 38-42). The diagnostic neural nets are necessary because neither McSnow nor the ML-based model make any a priori assumptions about particle size distributions. Hence, all bulk particle properties have to be learned from the training data.

All these networks are simple dense fully-connected multilayer perceptrons and the size for each network is specified in the Tables. The size has been determined by hyperparameter study and subsequent testing with the 1d and 2d simulations that are described in the following sections. The network size is relevant for the computation time needed and large neural nets would make the scheme considerably more expensive. This is especially important for processes that occur almost everywhere in a cloud, like depositional growth, and less so for very special processes that occur rarely like melting of graupel.

All 42 neural nets are regression models and directly provide the required physical variable (process rate, sedimentation velocity, etc.). This proved to work well in this case for all processes except for the self-collection of unrimed snow. Specifically for this process, we found that a two-step approach as described by Gettelman et al. (2021) is indeed beneficial and improves the performance of the overall scheme. The two-step approach uses a classifier network, which first determines whether the process rate is non-zero, followed by a regression network that estimates the actual process rate. Network no. 43 estimates the probability of self-collection of unrimed snow. Only where this probability exceeds 0.5 the regression network no. 7 is applied to calculate the corresponding process rates.

For all 55 process rates and the additional variables like sedimentation velocities, radar reflectivities, and effective radii, training data needs to be calculated from the McSnow output. This is straightforward and involves only sums over super-particles. Hence, no additional assumptions, choices, or approximations are necessary at this stage. For a derivation, we refer to section 3 of SR20, who detail this step for the warm-rain processes.

To generate a broad range of training samples, McSnow simulations are done for various atmospheric and microphysical conditions. Hence, we change the assumptions in the atmospheric profiles like domain height, height and thickness of the liquid layer, initial ice water content and ice number density of the parcel, the liquid water content and mean radius of the cloud droplets in the liquid layer, and the ice supersaturation outside of the liquid layer. Details are given in Table 3. This constitutes a multidimensional gridded hypercube from which we draw random samples. We did not perform a Latin hypercube sampling but drew samples in each box, which could be described as full hypercube sampling. Overall we have performed more than 20000 McSnow simulations for this study. The number of training and testing/validation samples for each process is given in Tables 4-6. These are of order 10^6 training samples for common processes like depositional growth, but the number drops down to 10^5 for many riming processes, and some processes that are rare or occur only in thin layers like self-collection of rimed ice or ice multiplication have only a few thousand training samples.

For the training of the individual neural nets Tensorflow 2.1 has been used. We randomly select 70 % of the data for training, 15 % for testing during the training process, and 15 % for validation after the training process. This split of the data is done for each process independently. We choose the mean squared error (MSE) as loss function, ReLU activation, an initial learning rate of 1e-3, the Adam optimizer, and early stopping (with a patience parameter between 5 and 10 depending on the process, and restoration of the best weights). All those choices are fairly standard for such simple regression networks with Tensorflow, but they seem to work well in our case. The training, validation, and testing data are standardized using the mean and standard deviation to ensure that all features have zero mean and a standard deviation of one. Hence, we apply the transformation $\check{\zeta} = (\zeta - \bar{\zeta})/\sigma_\zeta$, where ζ is a feature vector with mean $\bar{\zeta}$ and standard deviation σ_ζ, and $\check{\zeta}$ is the standardized value of the feature.

For most process rates, especially the collision rates, all dependencies are learned from the data. For some thermodynamic processes, we decided not to learn linear dependencies that are well-known. For example, the deposition rate is a linear function of the supersaturation S_i. We do not have to infer that from data. Hence, this linear dependency is removed from the training data. This explains why NNs 1-5 have no dependency on water vapor or supersaturation. Similarly, for melting rates of the internal melting, i.e. the melting within one particle category that converts ice or rime mass to liquid mass, we assume a linear dependency on the temperature deviation from the melting point. The latter is an approximation but works better in this case, because the melting layers can be thin and are then not sufficiently sampled by the training data.

For most of the neural nets the mean absolute error (MAE) and mean squared error (MSE) become sufficiently small. A value below 0.1 for MAE and below 0.05 for MSE is already sufficient given the complexity of the problem and the uncertainty of the assumptions made in such schemes. For some processes that are complicated but lack training data, the errors remain larger. Some processes involving rimed ice show the largest uncertainty because rimed ice exists only in rather thin layers which limits the number of training samples. At the same time, rimed ice can change its properties quite strongly due to the variability in rime density. At least, for the most frequent and most important processes it is possible to have enough training data to achieve a good and robust approximation.

5 Comparison with super-particle reference

A first and already quite challenging test for the ML-based model is to reproduce the training data when the individual process rates are combined into a system of ordinary differential equations (ODEs). This means that the process rates trained independently in the previous section constitute an ODE system and should reproduce the bulk variables of the McSnow simulation when used in concert. It is far from trivial that this works, because it requires that the process rates are sufficiently well approximated in a large part of the phase space. Hence, although we use the training data or very similar simulations for validation, this is a meaningful test for the ML-based model. The four-equation warm-rain model investigated in SR20 partly failed this test, because although the results of the ML model were reasonable approximations of the reference model, they were inferior to a well-established analytic parameterization.

Figures 1 and 2 present the results of such an ODE test of the ML-based microphysics. Shown are vertical profiles of water content and the number density of the various hydrometeor categories. As explained earlier, the vertical profiles are equivalent to time series, because the parcel (box model) falls through a prescribed atmospheric profile. The parcel starts with only unrimed cloud ice as initial condition. Soon unrimed snow forms by aggregation and starts to dominate the ice water content at 5000 m height, at 3500 m almost all cloud ice has been depleted. The unrimed snow reaches the liquid water layer at 3000 m height and almost instantaneously becomes rimed snow. Small amounts of unrimed ice and rimed ice exist within the liquid water zone. Unrimed ice can exist there as long the particles are smaller than the riming onset of approximately 100 μm. At 1500 m the particles reach the 0 °C level and the rimed snow melts into raindrops. The ML-based model can reproduce this archetypical behavior very well and the match with the original McSnow data is very good. The only small error is that the ML-based model is not able to produce sufficient amounts of unrimed and rimed ice in the liquid water zone. Note that the dotted lines, which represent the ODE system with the bulk process rates that would serve as training data (if this simulation would actually be part of the training data, which it is not), match the McSnow output even better and do capture the unrimed and rimed ice in the liquid layer. Hence, this information is in principle contained in the training data. The profiles of the number densities are much more complicated and show larger differences between McSnow and the ML-based bulk model. Nevertheless, the qualitative behavior is captured well by the ML-based model. The number densities show that the ML-based model does have some unrimed and rimed ice in the liquid water zone, but it is a factor 2-3 too low compared to McSnow. The profiles of the number densities also reveal some approximations that we made in the formulation of the bulk process rates at and below the melting level, and, hence, neither the training data (dotted lines) nor the ML-based model (dashed lines) match the reference of McSnow (solid lines) perfectly for melting particles and raindrops. Figure 3 shows some of the bulk particle properties of rimed snow for the same simulation as Fig. 1. Here the rime fraction matches quite well between McSnow and the ML-based bulk scheme, but the rime density of rimed snow shows initially a too rapid increase but then it flattens off and does not reach the same values as the reference. Melting happens in a rather thin layer of only 500 m, but at least the ML-based model does show a reasonable increase in melt fraction of rimed snow and, hence, captures the thickness of the melting layer quite well. Overall, the ML-based is able to pass this first test, because it provides a reasonable evolution of the microphysical variables including the prognostic particle properties like rime fraction and melt fraction for this archetypical but highly idealized case.

6 Squall line simulation with ICON

To perform idealized squall line simulations with the new ML-based P3-like microphysics, the scheme has been implemented in the ICON model (Zängl et al., 2015). For the ML-based scheme, the neural nets need to be evaluated in ICON as part of the model physics. To achieve this, the coefficients of all neural nets are stored in NetCDF files, which can easily be read into ICON. The evaluation of the neural nets, often called inference, is done using Fortran code originally developed for an ML-based satellite forward operator (Scheck, 2021). Other possible coupling strategies for using machine learning in ICON are for example discussed in Arnold et al. (2023).

To improve the efficiency of the implementation, especially on the NEC Aurora vector architecture currently in use at DWD, index lists are generated for each process. The index lists collect the grid points at which the input variables relevant for that process are non-zero, and the neural nets are then only evaluated where a non-zero process rate can be expected. This improves the efficiency not only on vector machines because many processes are non-zero only in very small parts of the three-dimensional domain, e.g., in deep convective updrafts with supercooled liquid water in case of riming processes. With this implementation the computational effort is bearable, but the scheme is considerably slower than the SB2006 two-moment microphysics. To some extent, because it has more prognostic variables (23 compared to 13), but the most expensive part of the ML-based scheme is in fact the inference of the neural nets. Some more implementation details are given in Appendix B.

To simulate a 3D idealized squall line the sounding of Weisman and Klemp (1982) is used with a linear wind profile from the surface to 2500 height and constant wind speed of 10 m/s above similar to Rotunno et al. (1988). The water vapor mixing ratio near the surface is 13 g/kg. The ICON simulation applies an R2B13 triangular icosahedral grid corresponding to an equivalent grid spacing of 308 m and a limited-area domain of 1.5 degree × 6.0 degree in the horizontal. The vertical grid has 128 levels with a domain top at 23 km and a damping layer starting at 20 km height. The TKE-based Mellor-Yamada level 2.5 boundary layer scheme is applied for vertical diffusion in combination with a 2D Smagorinsky closure in the horizontal. The Phillips et al. (2008) ice nucleation parameterization is used with constant number densities for dust, soot and organics given as $n_{\mathrm{dust}} = 1.6 \times 10^6\ \mathrm{m}^{-3}$, $n_{\mathrm{soot}} = 25 \times 10^6\ \mathrm{m}^{-3}$ and $n_{\mathrm{orga}} = 30 \times 10^6\ \mathrm{m}^{-3}$ similar to the 'high IN' setting of Seifert et al. (2012). The CCN activation is parameterized based on Segal and Khain (2006) with $N_{\mathrm{CN}} = 500 \times 10^6\ \mathrm{m}^{-3}$. In the following ICON simulations with the bulk two-moment scheme of Seifert and Beheng (2006, SB hereafter) are compared with the new ML-based P3-like bulk microphysics schemes.

The spatial structure of the squall line can be quantified with help of the radar reflectivity factor (dBZ). Observations often show a bimodal structure with high dBZ values in the convective core and a secondary weaker maximum in the trailing stratiform regions (see e.g. Figure 3 of Xue et al. (2017)). The separation of these two regions with a clear minimum in between, is difficult to capture with atmospheric models as discussed by Morrison et al. (2009) and Xue et al. (2017). Figure 4 shows vertical cross-sections of the radar reflectivity factor for ICON simulations with the SB scheme and the ML-based P3-like scheme. Using the SB scheme results in a relatively narrow squall line, which is dominated by the convective core and has no clear separation in convective and stratiform region. This is different for the ML-based P3-like scheme, which supports a more extended stratiform region with a more pronounced secondary maximum. Both microphysical schemes provide a reasonable squall line structure, but the ML-based scheme can alleviate some of the deficiencies of the SB schemes.

To achieve this improved spatial structure the ML-based scheme needs to be able to predict the evolution of the physical properties of the hydrometeors in the squall line. That the ML-based scheme is able to do this, is shown in Figure 5. The bulk rime fraction shows high values within the convective core where heavy riming occurs, the rime fraction decreases continuously within the stratiform region. This is reasonable, because little riming should happen outside the convective core, and particles with higher rime fraction have higher fall velocity and fall out more quickly. Decomposing the rime fraction in snow and graupel categories reveals that the rime fraction within the convective core is dominated by graupel, which has a rime fraction larger than 0.8. The stratiform region is almost only rimed snow and the rime fraction of rimed snow shows a maximum just behind the convective core. From there it decreases continuously because strongly rimed particles are removed by sedimentation. A similar structure is seen in rime density and the explanation is similar in the sense that riming is happening in the convective core and particles with high rime density have higher fall speeds and are removed by sedimentation. This suggests that the ML-based P3-like scheme does in fact capture the main physical processes and dependencies correctly. A more detailed analysis would require validation with in-situ observations or polarimetric radar data, which is beyond the scope of the current study.

Another interesting feature of the ML-based P3-like scheme is the explicit liquid mass of the rimed particle categories. For the squall line case, the liquid water fraction of rimed snow and graupel is shown in Figure 6. For rimed snow, the melting layer is roughly 2 km deep near the convective core and becomes thinner in the stratiform regions. This is again easily understood due to the larger and more heavily rimed particles closer to the convective core. Graupel reaches the ground in the convective core with a liquid fraction of 0.5. For graupel, the ML-based model predicts wet graupel up to 6 km height. This is physically possible because in zones of high supercooled liquid water, the riming rate becomes so large that the latent heat of freezing can no longer be dissipated by diffusion. This regime is called wet growth and is usually not represented in bulk microphysics schemes. The marginal liquid ratios between 0.1 and 0.01 below 8 km height are due to pockets of supercooled liquid water that can occur locally. Based on the results above, we can conclude that the ML-based P3-like scheme passed this idealized squall line test and delivers the improvement that can be expected from the P3 approach.

7 Mesoscale simulation with ICON

Machine learning models that are trained on simulation data may work well in idealized simulations as the squall line of the previous section, but can nevertheless fail when applied in a real-world situation. Hence, the next test is an example of an actual numerical weather prediction case using the ICON-D2 configuration of Deutscher Wetterdienst (DWD) similar to the operational regional forecast. The operational NWP system at DWD consists of a global ICON model, currently at 13 km grid spacing, with a European two-way nest at 6.5 km grid spacing (called ICON-EU), and the regional ICON-

D2 with approximately 2 km grid spacing over central Europe (Reinert et al., 2023). ICON uses an icosahedral unstructured mesh, which for ICON-D2 has 542040 cells on each of the 65 model levels, and a vertically stretched grid. For ICON-D2 the vertical grid spacing in the lowest levels is smaller than 100 m, but near the tropopause, it is approximately 500 m (Reinert et al., 2023). The domain top is at 22 km height. Operationally ICON-D2 still uses a one-moment microphysics, but the pre-operational rapid update cycle (RUC) applies the SB two-moment microphysics. The RUC spins off from the one-moment analysis at 0 UTC and performs its own analysis with the two-moment scheme using a local ensemble transform Kalman filter (LETKF, Schraff et al. (2016); Vobig et al. (2021)). Here we use the RUC analysis from 12 UTC to initialize forecasts for the afternoon of 19 May 2022. Note that the analysis has only been done with the SB two-moment scheme, not with the new ML-based P3-like scheme. The latter is currently not possible, because it would require coupling the ML-based scheme with the radar forward operator EMVORADO (Zeng et al., 2016), which is beyond the scope of the current study. Forecasts are performed for 6 hours and compared with the European Opera radar reflectivity composite. We use a radar reflectivity factor in Rayleigh approximation for both microphysics schemes to allow a fair comparison. For the ML-based microphysics scheme the neural nets 33-37 provide the second mass moment of the particle size distribution, which is required for the radar reflectivity factor in Rayleigh approximation.

Figure 7 presents the column maximum radar reflectivity for 13:30 UTC. Note that the ICON-D2 domain used for these simulations is considerably larger than the area shown in the Figure. The Opera composite shows a squall line over the Netherlands and Belgium approaching Germany. The southern end of the line shows a narrow convective region with high reflectivity values, to the north a larger stratiform region is visible. The ICON-D2 forecast with the SB two-moment scheme captures the overall structure of the convective system, but the convective cores are too weak and the stratiform region is too narrow and not as extended as in the observations. These are typical biases of ICON-D2 with the SB two-moment scheme, which are quite pronounced in this case. With the ML-based P3-like scheme the convective line at the southern end of the convective complex is even weaker, although higher reflectivity values occur within active convective cores. The stratiform region is more extended compared to SB but is more symmetric around the convective line and does not resemble the observations better than the simulation using SB microphysics. Hence, in contrast to the idealized squall line, the ML-based P3-like scheme does not improve over the SB scheme in this real-case application. The improved structure in the idealized squall line simulation is only apparent at the high spatial resolution of the 308 m mesh and deteriorates on coarser grids (not shown). Increasing the resolution of the real-case simulation to a sub-km mesh would unfortunately be too costly with the ML-based microphysics as it is currently implemented. Hence, the result of this case study is undecisive. The ML-based P3-like scheme is stable and provides a reasonable representation of the mesoscale convective system, but it can not improve over the SB two-moment scheme in this case. A simple explanation for this result is that the microphysics scheme is not the limiting factor for the forecast quality in this case. It is very likely that the model dynamics and especially the boundary layer scheme play an important role and contribute to the deficiencies of this ICON forecast.

8 Summary and Conclusions

Machine learning has been applied to build a complex bulk microphysics scheme, which predicts not only particle mass and number but detailed physical properties like rime mass, rime volume, and liquid mass following the P3 approach of Morrison and Milbrandt (2015). Training data has been generated using idealized simulations with the super-particle model McSnow. Hence, the machine learning performs a coarse-graining of the detailed McSnow data to a bulk microphysics scheme. The human role in this process is twofold: First, to make an a priori choice of the prognostic equations, i.e. the number of particle categories and the prognostic variables for each category. Second, to design the McSnow simulations that provide the training data. Based on these two preparatory steps, the machine learning workflow is almost automatic and does not require much human intervention, except for some limited hyperparameter tuning. Standard regression neural nets are sufficient for most processes. Only for the self-collection of unrimed snow, we found that a two-step classifier-regression approach as recommended by Gettelman et al. (2021) is superior to using only a regression neural net. The ML-based P3-like microphysics scheme has been implemented in the ICON weather and climate model using Fortran code for the inference of fully connected neural nets.

The ML-based P3-like microphysics scheme has passed three relevant tests: First, it can reproduce simulations similar to the training data, which requires that the individual process rates work in concert to reproduce the behavior of McSnow in an ODE sense. This is by no means trivial as shown for example by SR20 for warm-rain microphysics. Second, the ML-based P3-like microphysics provides reasonable results for a 3d idealized squall line simulation with ICON. It can in some aspects improve over the SB two-moment scheme in that it produces a more realistic-looking extended stratiform region with a secondary maximum in radar reflectivity. The ML-based scheme achieves this by predicting physically plausible rime mass and rime density of snow and graupel and corresponding sedimentation velocities. The ML-based P3-like scheme is even able to predict the wet growth regime of graupel within the convective core. Third, the ML-based scheme has been applied in a realistic forecast scenario with ICON on a 2 km grid to predict the evolution of a mesoscale convective system. In this case, the ML-based scheme runs stably over a large spatial domain and provides a reasonable representation of the cloud microphysics. Unfortunately, it is not able to improve over the SB two-moment scheme in the chosen case. Most likely, the microphysics is simply not the limiting factor for the forecast quality of this mesoscale convective system, but instead, other model components are relevant as well and would have to be improved.

In contrast to classic bulk microphysics parameterizations, the ML-based P3-like scheme does not explicitly make assumptions regarding particle geometries or particle size distributions. All this is learned from the McSnow simulations in a parameter-free way. This makes the ML approach flexible, but it requires that additional neural nets are trained for diagnostics like radar reflectivity or effective radius. For more complex diagnostics like polarimetric radar variables, which are very challenging for conventional bulk microphysics schemes, the ML approach could be promising, though.

The ML approach chosen here has some disadvantages. First, the ML-based scheme ended up being computationally expensive. On one hand, simply because we decided to build a very complicated scheme with 23 prognostic variables. On the other hand, the implementation with individual neural nets for each physical process that have to be evaluated at each grid point and each time step is rather inefficient. It should be possible to subsequently build an emulator of the ML-based scheme that would overcome these deficiencies. For example, by having fewer neural networks and taking model columns as input instead of individual grid points. Even the calculation of process tendencies can be questioned and instead, a direct mapping of state variables from one time step to the next could be implemented.

Second, the training data is based on rather simplistic box model simulations, which does not make full use of the super-particle model McSnow. McSnow can in principle be applied in 2d or even 3d ICON simulations. Hence, the current ML-based scheme could be used as a baseline model and more training data from 2d and 3d McSnow simulations could further improve the realism of the microphysical processes and their interaction.

A third and more general issue of such ML-based schemes is that they do not allow much a posteriori tuning of the model. Any change in the microphysical assumptions like basic particle geometry or sticking or collision efficiencies, for example, would have to be done in McSnow. Then the full ML workflow has to be repeated including the production of the training data. This makes sensitivity studies to explore model uncertainties very time-consuming. In practice, NWP or climate models do require some a posteriori tuning to balance different physical processes and their biases. Common tuning parameters like intercept parameters of the particle size distribution, terminal fall velocity, or particle geometries cannot easily be modified in ML-based schemes. Changing the bulk sedimentation velocity by a constant factor is possible, but would make the scheme inconsistent. This issue could only be overcome if the ML-based model could be further trained and improved within the atmospheric model itself. Preferably such an online training would be done with actual observations as part of a data assimilation system. First steps toward such an online training capability for ICON are currently being implemented at DWD.

Figure 1. Vertical profiles of mass densities of the various particle classes of the ML-based P3-like scheme. Shown is McSnow output (solid), the ODE solution using training data (dotted) and the ODE solution of the ML-based scheme (dashed). Shown is a simulation with $h_1 = 1500$ m, $h_2 = 3000$ m, $r_c = 15$ μm, $N_i = 20$ dm^{-3}, $Q_i = 0.2$ g cm^{-3}, $Q_c = 0.1$ g cm^{-3} and $S_i = 0$.

Figure 2. As Figure 1, but for number densities.

Figure 3. Vertical profiles of particle properties of rimed snow. Shown is McSnow output (solid), the ODE solution using training data (dotted), and the ODE solution of the ML-based scheme (dashed).

Figure 4. Vertical cross section of radar reflectivity dBZ for the SB two-moment scheme (left) and the ML-based P3-like scheme (right). Shown are averages along the y-direction after 300 min simulation time.

Figure 5. Vertical cross sections of rime fraction (left) and rime density (right), of all hydrometeors (top), rimed snow (center) and graupel (bottom) as predicted by the ML-based P3-like scheme. Shown are averages along the y-direction after 300 min simulation time.

Figure 6. Vertical cross section of the liquid water fraction of rimed snow (left) and graupel (right) of the ML-based P3-like scheme. Shown are averages along the y-direction after 300 min simulation time.

Figure 7. Column maximum radar reflectivity for 19 May 2022, 13:30 UTC for a central European region. Shown is the Opera composite (left), the ICON-D2 simulation using the SB two-moment microphysics (center) and ICON using the ML-based P3-like two-moment scheme (right).

Table 4. Overview of ML models of ice microphysical processes. Mean absolute error (MAE) and mean squared error (MSE) are shown for the normalized testing data and are dimensionless. All ML models use ReLU activation and apply the Adam optimizer.

No.	physical process	predictors (features)	output (labels)	NN size	training samples	testing samples	MAE	MSE
1	deposition/sublimation of unrimed ice	q_i, n_i, T, ρ	$\partial_t q_i/S_i$	8x1	1,234,870	263,676	0.0130	0.0036
2	deposition/sublimation of unrimed snow	q_s, n_s, T, ρ	$\partial_t q_s/S_i$	6x1	1,279,759	273,979	0.0307	0.0059
3	deposition/sublimation of graupel	q_g, n_g, ψ_g, ϕ_g, T, ρ	$\partial_t q_g/S_i$	6x1	464,361	99,118	0.0491	0.0253
4	deposition/sublimation of rimed ice	$q_{\rm ri}$, $n_{\rm ri}$, $\psi_{\rm ri}$, $\phi_{\rm ri}$, T, ρ	$\partial_t q_{\rm ri}/S_i$	6x1	66,490	14,288	0.139	0.0987
5	deposition/sublimation of rimed snow	$q_{\rm rs}$, $n_{\rm rs}$, $\psi_{\rm rs}$, $\phi_{\rm rs}$, T, ρ	$\partial_t q_{\rm rs}/S_i$	6x1	263,859	57,038	0.136	0.130
6	selfcollection of unrimed ice	q_i, n_i, T, ρ	$\partial_t q_i$, $\partial_t n_i$	16x2	1,212,134	259,225	0.256	0.124
7	selfcollection of unrimed snow	q_s, n_s, T, ρ	$\partial_t n_i$	32x2	1,135,918	243,714	0.104	0.023
8	aggregation of unrimed ice and unrimed snow	q_i, n_i, q_s, n_s, T, ρ	$\partial_t n_i$, $\partial_t q_i$	16x2	1,497,110	320,229	0.073	0.0178
9	selfcollection of rimed ice	$q_{\rm ri}$, $n_{\rm ri}$, $\psi_{\rm ri}$, $\phi_{\rm ri}$, T, ρ	$\partial_t n_{\rm ri}$, $\partial_t q_{\rm ri}$, $\partial_t \psi_{\rm ri}$, $\partial_t \phi_{\rm ri}$	16x2	13,759	2,971	0.307	0.173
10	selfcollection of rimed snow	$q_{\rm rs}$, $n_{\rm rs}$, $\psi_{\rm rs}$, $\phi_{\rm rs}$, $\hat{\ell}_{\rm rs}$	$\partial_t n_{\rm rs}$	16x2	1,425,976	305,885	0.0829	0.0177
11	aggregation of unrimed ice and rimed snow	q_i, n_i, $q_{\rm rs}$, $n_{\rm rs}$, $\psi_{\rm rs}$, $\phi_{\rm rs}$, T, ρ	$\partial_t n_i$, $\partial_t q_i$	16x2	337,354	72,545	0.0688	0.0170
12	aggregation of unrimed snow and rimed snow	q_s, n_s, $q_{\rm rs}$, $n_{\rm rs}$, $\psi_{\rm rs}$, $\phi_{\rm rs}$, T, ρ	$\partial_t n_s$, $\partial_t q_s$	16x2	238,671	51,669	0.0823	0.0192
13	aggregation of rimed ice and rimed snow	$q_{\rm ri}$, $n_{\rm ri}$, $\psi_{\rm ri}$, $\phi_{\rm ri}$, $q_{\rm rs}$, $n_{\rm rs}$, $\psi_{\rm rs}$, $\phi_{\rm rs}$, T, ρ	$\partial_t n_{\rm ri}$, $\partial_t q_{\rm ri}$, $\partial_t \psi_{\rm ri}$, $\partial_t \phi_{\rm ri}$	16x2	111,188	24,006	0.277	0.180

Table 5. Overview of ML models (continued)

No.	physical process	predictors (features)	output (labels)	NN size	training samples	testing samples	MAE	MSE
14	conversion of ice to rimed ice	q_i, n_i, q_c, r_c, T, ρ	$\partial_t n_{\mathrm{ri}}$, $\partial_t q_{\mathrm{ri}}$, $\partial_t \psi_{\mathrm{ri}}$, $\partial_t \phi_{\mathrm{ri}}$	16x2	12,059	2,495	0.236	0.236
15	conversion of snow to rimed snow	q_s, n_s, q_c, r_c, T, ρ	$\partial_t n_{\mathrm{rs}}$, $\partial_t q_{\mathrm{rs}}$, $\partial_t \psi_{\mathrm{rs}}$, $\partial_t \phi_{\mathrm{rs}}$	16x2	10,492	2,245	0.206	0.102
16	riming of rimed ice	q_{ri}, n_{ri}, ψ_{ri}, ϕ_{ri}, q_c, r_c, T, ρ	$\partial_t q_{\mathrm{ri}}$, $\partial_t \psi_{\mathrm{ri}}$, $\partial_t \phi_{\mathrm{ri}}$	16x2	33,505	7,141	0.152	0.0753
17	riming of rimed snow with cloud droplets	q_{rs}, n_{rs}, ψ_{rs}, ϕ_{rs}, $\hat{\ell}_{\mathrm{rs}}$, q_c, r_c, T	$\partial_t q_{\mathrm{rs}}$, $\partial_t n_{\mathrm{rs}}$, $\partial_t \psi_{\mathrm{rs}}$, $\partial_t \phi_{\mathrm{rs}}$, $\partial_t \ell_{\mathrm{rs}}$	16x2	227,928	49,345	0.114	0.0665
18	riming of graupel with cloud droplets	q_g, n_g, ψ_g, ϕ_g, $\hat{\ell}_g$, q_c, r_c, T, ρ	$\partial_t q_g$, $\partial_t n_g$, $\partial_t \psi_g$, $\partial_t \phi_g$, $\partial_t \ell_g$	16x2	530,385	113,628	0.0966	0.0481
19	internal melting of rimed snow	q_{rs}, n_{rs}, ψ_{rs}, ϕ_{rs}, $\hat{\ell}_{\mathrm{rs}}$	$\partial_t \ell_{\mathrm{rs}}/(T - T_3)$	16x2	810,109	173,193	0.0227	0.0027
20	melting of rimed snow to rain	q_{rs}, n_{rs}, ψ_{rs}, ϕ_{rs}, $\hat{\ell}_{\mathrm{rs}}$, T	$\partial_t \ell_{\mathrm{rs}}$, $\partial_t n_{\mathrm{rs}}$	16x2	418,529	89,765	0.179	0.0901
21	internal melting of graupel	q_g, n_g, ψ_g, ϕ_g, $\hat{\ell}_g$	$\partial_t \ell_g/(T - T_3)$	16x2	558,006	119,939	0.020	0.0020
22	melting of graupel to rain	q_g, n_g, ψ_g, ϕ_g, $\hat{\ell}_g$, T	$\partial_t \ell_g$,$\partial_t n_g$	16x2	311,931	66,764	0.154	0.088
23	internal freezing of graupel	q_g, n_g, ψ_g, ϕ_g, $\hat{\ell}_g$, q_c, r_c, q_r, n_r	$\partial_t \ell_g/(T_3 - T)$	16x2	420,439	90,226	0.151	0.088
24	collection of rain by graupel	q_g, n_g, ψ_g, ϕ_g, $\hat{\ell}_g$, q_r, n_r, T, ρ	$\partial_t \ell_g$	16x2	671,660	144,110	0.124	0.0624
25	ice multiplication for graupel	q_g, n_g, ψ_g, ϕ_g, $\hat{\ell}_g$, T, ρ	$\partial_t n_i$	16x2	118,906	25,574	0.0894	0.0272
26	ice multiplication for rimed snow	q_{rs}, n_{rs}, ψ_{rs}, ϕ_{rs}, $\hat{\ell}_{\mathrm{rs}}$, T, ρ	$\partial_t n_i$	16x2	68,062	14,673	0.120	0.0467
27	ice multiplication for rimed ice	q_{ri}, n_{ri}, ψ_{ri}, ϕ_{ri}, T, ρ	$\partial_t n_i$	16x2	9,819	2,068	0.109	0.0330
28	sedimentation of unrimed ice	q_i, n_i, T, ρ	v_q, v_n	16x2	979,504	209,552	0.078	0.0163

Table 6. Overview of ML models (continued)

No.	physical process	predictors (features)	output (labels)	NN size	training samples	testing samples	MAE	MSE
29	sedimentation of unrimed snow	q_s, n_s, T, ρ	v_q, v_n	16x1	1,707,255	365,700	0.076	0.020
30	sedimentation of rimed ice	q_{ri}, n_{ri}, ψ_{ri}, ϕ_{ri}, T, ρ	v_q, v_n, v_ψ	32x2	152,090	32,373	0.0863	0.0177
31	sedimentation of rimed snow	q_{rs}, n_{rs}, ψ_{rs}, ϕ_{rs}, $\hat{\ell}_{\mathrm{rs}}$, T, ρ	v_q, v_n, v_ψ, v_ℓ	32x2	1,500,619	321,431	0.0732	0.0194
32	sedimentation of graupel	q_g, n_g, ψ_g, ϕ_g, $\hat{\ell}_g$, ρ	v_q, v_n, v_ψ, v_ℓ	32x2	1,859,964	398,612	0.0562	0.0113
33	radar reflectivity of graupel	q_g, n_g, ψ_g, ϕ_g	z_g	16x2	416,676	89,544	0.0713	0.0118
34	radar reflectivity of rimed snow	q_{rs}, n_{rs}, ψ_{rs}, ϕ_{rs}	z_{rs}	16x2	1,076,404	230,621	0.084	0.017
35	radar reflectivity of rimed ice	q_{ri}, n_{ri}, ψ_{ri}, ϕ_{ri}	z_{ri}	16x2	161,476	34,489	0.051	0.0051
36	radar reflectivity of unrimed ice	q_i, n_i, T, ρ	z_i	16x2	1,122,035	240,143	0.045	0.0065
37	radar reflectivity of unrimed snow	q_s, n_s, T, ρ	z_s	16x2	1,158,769	248,745	0.049	0.0075
38	effective radius of unrimed ice	q_i, n_i, T	$r_{eff,i}$	16x2	1,666,618	356,002	0.0460	0.0084
39	effective radius of unrimed snow	q_s, n_s, T	$r_{eff,s}$	16x2	1,120,739	240,259	0.0468	0.0084
40	effective radius of rimed ice	q_{ri}, n_{ri}, ψ_{ri}, ϕ_{ri}	$r_{eff,\mathrm{ri}}$	16x2	182,967	39,466	0.1409	0.0950
41	effective radius of rimed snow	q_{rs}, n_{rs}, ψ_{rs}, ϕ_{rs}, $\hat{\ell}_{\mathrm{rs}}$	$r_{eff,\mathrm{rs}}$	16x2	1,497,472	320,385	0.0519	0.0323
42	effective radius of graupel	q_g, n_g, ψ_g, ϕ_g, $\hat{\ell}_g$	$r_{eff,g}$	16x2	1,861,272	398,483	0.0381	0.0063
	classifier network						accuracy	
43	selfcollection of unrimed snow	q_s, n_s, n_i, T, ρ	P_{self}	16x2	1,719,206	368,237	0.938	

Appendix A Sticking efficiency in McSnow

There are some observations and laboratory measurements of the sticking efficiency of ice crystals as a function of temperature (Hosler & Hallgren, 1960; Mitchell, 1988; Kajikawa & Heymsfield, 1989; Connolly et al., 2012) but for graupel-graupel collisions or partially rimed snowflakes, the authors are not aware of any measurements. Phillips et al. (2015) discuss the dependency of the sticking efficiency on the collision kinetic energy and, hence, provide a theoretical framework to explain the decrease of the sticking efficiency with increasing degree of riming. Due to the lack of data, a consistent physically-based parameterization is beyond the scope of this study. Nevertheless, a reasonable and continuous description is required to generate meaningful training data for the P3 approach. In the current study, we use the degree of riming defined as

$$\xi = \frac{m_r + m_f}{m_r + m_f + m_i + m_\ell} = \frac{m_r + m_f}{m_{\mathrm{tot}}} \tag{A1}$$

with the rime mass m_r, the frozen mass m_f, the ice (crystal) mass m_i, the liquid mass m_ℓ and the total particle mass m_{tot}.

The following ad-hoc parameterization for the sticking efficiency E_s of two particles a and b has been used in the McSnow simulations:

$$E_s = \begin{cases} E_i, & \text{for } \xi_a + \xi_b < \xi_1 \\ E_g, & \text{for } \xi_a + \xi_b > \xi_2 \\ E_i \frac{\xi_a + \xi_b - \xi_2}{\xi_1 - \xi_2} + E_g \frac{\xi_a + \xi_b - \xi_1}{\xi_2 - \xi_1} & \end{cases} \tag{A2}$$

with $\xi_1 = 0.01$ and $\xi_2 = 0.9$. Here E_i is the temperature-dependent piecewise linear sticking efficiency of unrimed crystals

$$E_i = \begin{cases} 0.07, & \text{for } T_c \geq 0\,^\circ\mathrm{C} \\ -0.005\,(T_c + 10) + 0.12, & \text{for } 0\,^\circ\mathrm{C} > T_c \geq -10\,^\circ\mathrm{C} \\ -0.040\,(T_c + 15) + 0.32, & \text{for } -10\,^\circ\mathrm{C} > T_c \geq -15\,^\circ\mathrm{C} \\ 0.050\,(T_c + 20) + 0.14, & \text{for } -15\,^\circ\mathrm{C} > T_c \geq -20\,^\circ\mathrm{C} \\ 0.0025\,(T_c + 40) + 0.04, & \text{for } -20\,^\circ\mathrm{C} > T_c \geq -40\,^\circ\mathrm{C} \\ 0.02, & \text{for } -40\,^\circ\mathrm{C} > T_c \end{cases} \tag{A3}$$

where T_c is the temperature in degrees Celsius. This formula is largely based on data of Connolly et al. (2012) as shown by their Figure 14, but we intentionally decided on the lower range of those measurements. This choice leads to a more pronounced trailing stratiform region of the idealized squall line.

$$E_g = 0.01 \tag{A4}$$

is the sticking efficiency of graupel-graupel collisions. This sticking efficiency is only applied to ice particles that have no liquid water at the particle surface. For melting particles and in the wet growth regime the sticking efficiency is set to one in McSnow. Both, E_i and E_g, are often used as tuning parameters in cloud simulations. See, for example, the discussion in Karrer et al. (2021) for the sticking efficiency of aggregates.

Appendix B Some implementation details in ICON

To be able to run the ML-based P3-like scheme stably in ICON, a few constraints are necessary. The values of cloud liquid water content and mean cloud particle radius are limited to a range not far beyond the training data. For cloud liquid water this is only an upper bound of 20×10^{-3} kg m^{-3}, which should rarely be reached. The cloud droplet radius is forced to be within 5-30 μm when passed to the neural networks. For the sedimentation velocities, upper and lower limits are imposed as always in the SB two-moment scheme. In addition, it is enforced that the sedimentation velocity of mass is larger than that for number. This is done by simply using the larger of the two velocities provided by the neural net for mass, whereas the smaller one is used for number. For the sink term of cloud droplet or raindrop number by riming the mean mass is assumed to be constant as the NNs do not yet provide the information about the size of the collected liquid drops. All microphysical processes except diffusional growth (deposition/sublimation) are only calculated if the mass content of the hydrometeor class exceeds 1×10^{-9} kg m^{-3}. For self-collection of rimed snow lower limits of 1×10^{-5} kg m^{-3} and 10 m^{-3} have to be exceeded for mass and number density, respectively. Melting is only calculated a long as the liquid fraction is below 0.99, then the remaining mass is instantly converted to rain. Conversion of rimed snow to rain by melting only happens if the bulk liquid fraction exceeds 0.3. All processes that are only relevant in the mixed-phase regime, are only calculated in the temperature range 236 K to 273 K. Ice multiplication is restricted to 265.9 K and 270.1 K and limited to 100 m^{-3} s^{-1}. The ML-based P3-like scheme is implemented in ICON as an extension of the SB two-moment scheme. The SB two-moment microphysics additionally enforces the particle sizes of each hydrometeor class to be within a physically meaningful range.

References

Aligo, E. A., Ferrier, B., & Carley, J. R. (2018). Modified NAM microphysics for forecasts of deep convective storms. *Mon. Wea. Rev.*, *146*(12), 4115–4153.

Arnold, C., Sharma, S., Weigel, T., & Greenberg, D. (2023). Efficient and Stable Coupling of the SuperdropNet Deep Learning-based Cloud Microphysics (v0.1.0) to the ICON Climate and Weather Model (v2.6.5). *EGUsphere*, *2023*, 1–17.

Barklie, R., & Gokhale, N. (1959). *The freezing of supercooled water drops.* (MW-30)

Bauer, P., Dueben, P. D., Hoefler, T., Quintino, T., Schulthess, T. C., & Wedi, N. P. (2021). The digital revolution of Earth-system science. *Nature Computational Science*, *1*(2), 104–113.

Bauer, P., Stevens, B., & Hazeleger, W. (2021). A digital twin of Earth for the green transition. *Nature Climate Change*, *11*(2), 80–83.

Böhm, J. P. (1992a). A general hydrodynamic theory for mixed-phase microphysics. Part I: drag and fall speed of hydrometeors. *Atmos. Res.*, *27*(4), 253-274.

Böhm, J. P. (1992b). A general hydrodynamic theory for mixed-phase microphysics. Part II: Collision kernels for coalescence. *Atmos. Res.*, *27*(4), 275–290.

Böhm, J. P. (1992c). A general hydrodynamic theory for mixed-phase microphysics. Part III: Riming and aggregation. *Atmos. Res.*, *28*(2), 103–123.

Böhm, J. P. (1994). Theoretical collision efficiencies for riming and aerosol impaction. *Atmos. Res.*, *32*(1-4), 171–187.

Böhm, J. P. (1999). Revision and clarification of "A general hydrodynamic theory for mixed-phase microphysics". *Atmos. Res.*, *52*(3), 167–176.

Böhm, J. P. (2004). Reply to Comment on "Revision and clarification of 'A general hydrodynamic theory for mixed-phase microphysics'". *Atmos. Res.*, *69*(3-4), 289–293.

Brdar, S., & Seifert, A. (2018). McSnow: A Monte-Carlo particle model for riming and aggregation of ice particles in a multidimensional microphysical phase space. *J. Adv. in Modeling Earth Systems*, *10*(1), 187-206. doi: https://doi.org/10.1002/2017MS001167

Bringi, V., Seifert, A., Wu, W., Thurai, M., Huang, G.-J., & Siewert, C. (2020). Hurricane dorian outer rain band observations and 1d particle model simulations: A case study. *Atmosphere*, *11*(8), 879. doi: 10.3390/atmos11080879

Chantry, M., Hatfield, S., Dueben, P., Polichtchouk, I., & Palmer, T. (2021). Machine learning emulation of gravity wave drag in numerical weather forecasting. *J. Adv. in Modeling Earth Systems*, *13*(7), e2021MS002477. doi: https://doi.org/10.1029/2021MS002477

Cholette, M., Milbrandt, J. A., Morrison, H., Paquin-Ricard, D., & Jacques, D. (2023). Combining triple-moment ice with prognostic liquid fraction in the P3 microphysics scheme: Impacts on a simulated squall line. *J. Adv. in Modeling Earth Systems*, *15*(4), e2022MS003328. (e2022MS003328 2022MS003328)

Cholette, M., Morrison, H., Milbrandt, J. A., & Thériault, J. M. (2019). Parameterization of the bulk liquid fraction on mixed-phase particles in the predicted particle properties (P3) scheme: Description and idealized simulations. *J. Atmos. Sci.*, *76*(2), 561 - 582. doi: 10.1175/JAS-D-18-0278.1

Cholette, M., Thériault, J. M., Milbrandt, J. A., & Morrison, H. (2020). Impacts of predicting the liquid fraction of mixed-phase particles on the simulation of an extreme freezing rain event: The 1998 north american ice storm. *Mon. Wea. Rev.*, *148*(9), 3799 - 3823. doi: 10.1175/MWR-D-20-0026.1

Cober, S. G., & List, R. (1993). Measurements of the heat and mass transfer parameters characterizing conical graupel growth. *J. Atmos. Sci.*, *50*(11), 1591–1609.

Connolly, P., Emersic, C., & Field, P. (2012). A laboratory investigation into the aggregation efficiency of small ice crystals. *Atmos. Chem. Phys.*, *12*(4), 2055–2076.

Feingold, G., Walko, R., Stevens, B., & Cotton, W. (1998). Simulations of marine stratocumulus using a new microphysical parameterization scheme. *Atmos. Res.*, *47*, 505–528.

Field, P. R., Lawson, R. P., Brown, P. R., Lloyd, G., Westbrook, C., Moisseev, D., … others (2017). Secondary ice production: Current state of the science and recommendations for the future. *Meteorological Monographs*, *58*, 7–1.

Gettelman, A., Gagne, D. J., Chen, C.-C., Christensen, M., Lebo, Z., Morrison, H., & Gantos, G. (2021). Machine learning the warm rain process. *J. Adv. in Modeling Earth Systems*, *13*(2), e2020MS002268.

Hallett, J., & Mossop, S. (1974). Production of secondary ice particles during the riming process. *Nature*, *249*(5452), 26–28.

Hosler, C., & Hallgren, R. (1960). The aggregation of small ice crystals. *Discussions of the Faraday Society*, *30*, 200–207.

Jakob, C. (2010). Accelerating progress in global atmospheric model development through improved parameterizations: Challenges, opportunities, and strategies. *Bull. Am. Met. Soc.*, *91*(7), 869–876.

Kajikawa, M., & Heymsfield, A. J. (1989). Aggregation of ice crystals in cirrus. *J. Atmos. Sci.*, *46*(20), 3108–3121.

Kärcher, B., Hendricks, J., & Lohmann, U. (2006). Physically based parameterization of cirrus cloud formation for use in global atmospheric models. *J. Geophys. Res.*, *111*(D1).

Karrer, M., Seifert, A., Ori, D., & Kneifel, S. (2021). Improving the representation of aggregation in a two-moment microphysical scheme with statistics of multi-frequency doppler radar observations. *Atmos. Chem. Phys.*, *21*(22), 17133–17166.

Khain, A., Beheng, K., Heymsfield, A., Korolev, A., Krichak, S., Levin, Z., ... Yano, J.-I. (2015). Representation of microphysical processes in cloud-resolving models: Spectral (bin) microphysics versus bulk parameterization. *Reviews of Geophysics*, *53*(2), 247–322.

Köhler, C. G., & Seifert, A. (2015). Identifying sensitivities for cirrus modelling using a two-moment two-mode bulk microphysics scheme. *Tellus B: Chemical and Physical Meteorology*, *67*(1), 24494.

Lagerquist, R., Turner, D., Ebert-Uphoff, I., Stewart, J., & Hagerty, V. (2021). Using deep learning to emulate and accelerate a radiative transfer model. *J. Atmos. Ocean. Tech.*, *38*(10), 1673–1696.

Meyer, D., Grimmond, S., Dueben, P., Hogan, R., & van Reeuwijk, M. (2022). Machine learning emulation of urban land surface processes. *J. Adv. in Modeling Earth Systems*, *14*(3), e2021MS002744. doi: https://doi.org/10.1029/2021MS002744

Meyer, D., Hogan, R. J., Dueben, P. D., & Mason, S. L. (2022). Machine learning emulation of 3d cloud radiative effects. *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002550. doi: https://doi.org/10.1029/2021MS002550

Milbrandt, J. A., & Morrison, H. (2016). Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part III: Introduction of multiple free categories. *J. Atmos. Sci.*, *73*(3), 975–995.

Milbrandt, J. A., Morrison, H., Dawson II, D. T., & Paukert, M. (2021). A triple-moment representation of ice in the predicted particle properties (P3) microphysics scheme. *J. Atmos. Sci.*, *78*(2), 439–458.

Mitchell, D. L. (1988). Evolution of snow-size spectra in cyclonic storms. Part I: Snow growth by vapor deposition and aggregation. *J. Atmos. Sci.*, *45*(22), 3431–3451.

Morrison, H., & Milbrandt, J. A. (2015). Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part I: Scheme description and idealized tests. *J. Atmos. Sci.*, *72*(1), 287–311.

Morrison, H., Thompson, G., & Tatarskii, V. (2009). Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Wea. Rev.*, *137*(3), 991 - 1007. doi: 10.1175/2008MWR2556.1

Mosimann, L., Weingartner, E., & Waldvogel, A. (1994). An analysis of accreted drop sizes and mass on rimed snow crystals. *J. Atmos. Sci.*, *51*(11), 1548 - 1558. doi: 10.1175/1520-0469(1994)051⟨1548:AAOADS⟩2.0.CO;2

Phillips, V. T., DeMott, P. J., & Andronache, C. (2008). An empirical parameterization of heterogeneous ice nucleation for multiple chemical species of aerosol. *J. Atmos. Sci.*, *65*(9), 2757–2783.

Phillips, V. T., Formenton, M., Bansemer, A., Kudzotsa, I., & Lienert, B. (2015). A parameterization of sticking efficiency for collisions of snow and graupel with ice crystals: Theory and comparison with observations. *J. Atmos. Sci.*, *72*(12), 4885–4902.

Rasmussen, R. M., & Heymsfield, A. J. (1987a). Melting and shedding of graupel and hail. Part II: Sensitivity study. *J. Atmos. Sci.*, *44*(19), 2764–2782.

Rasmussen, R. M., & Heymsfield, A. J. (1987b). Melting and shedding of graupel and hail. Part I: Model physics. *J. Atmos. Sci.*, *44*(19), 2754–2763.

Rasmussen, R. M., Levizzani, V., & Pruppacher, H. (1984a). A wind tunnel and theoretical study of the melting behavior of atmospheric ice particles. II: A theoretical study for frozen drops of radius < 500 μm. *J. Atmos. Sci.*, *41*(3), 374–380.

Rasmussen, R. M., Levizzani, V., & Pruppacher, H. (1984b). A wind tunnel and theoretical study on the melting behavior of atmospheric ice particles: III. Experiment and theory for spherical ice particles of radius $> 500\ \mu$m. *J. Atmos. Sci.*, *41*(3), 381–388.

Reinert, D., Prill, F., Frank, H., Denhard, M., Baldauf, M., Schraff, C., ... Zängl, G. (2023). *DWD database reference for the global and regional ICON and ICON-EPS forecasting system* (Technical Report and Database Description, Version 2.2.2). Deutscher Wetterdienst. Retrieved from `https://www.dwd.de/SharedDocs/downloads/DE/modelldokumentationen/nwv/icon/icon_dbbeschr_aktuell.html`

Rotunno, R., Klemp, J. B., & Weisman, M. L. (1988). A theory for strong, long-lived squall lines. *J. Atmos. Sci.*, *45*(3), 463–485.

Scheck, L. (2021). A neural network based forward operator for visible satellite images and its adjoint. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *274*, 107841.

Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Periáñez, A., & Potthast, R. (2016). Kilometre-scale ensemble data assimilation for the cosmo model (kenda). *Quart. J. Roy. Met. Soc.*, *142*(696), 1453–1472.

Segal, Y., & Khain, A. (2006). Dependence of droplet concentration on aerosol conditions in different cloud types: Application to droplet concentration parameterization of aerosol conditions. *J. Geophys. Res.*, *111*(D15).

Seifert, A. (2008). On the parameterization of evaporation of raindrops as simulated by a one-dimensional rainshaft model. *J. Atmos. Sci.*, *65*(11), 3608–3619.

Seifert, A., & Beheng, K. D. (2001). A double-moment parameterization for simulating autoconversion, accretion and selfcollection. *Atmos. Res.*, *59-60*, 265–281.

Seifert, A., & Beheng, K. D. (2006). A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 1: Model description. *Meteorol. Atmos. Phys.*, *92*(1-2), 45–66.

Seifert, A., Köhler, C., & Beheng, K. (2012). Aerosol-cloud-precipitation effects over germany as simulated by a convective-scale numerical weather prediction model. *Atmospheric Chemistry and Physics*, *12*(2), 709–725.

Seifert, A., Leinonen, J., Siewert, C., & Kneifel, S. (2019). The geometry of rimed aggregate snowflakes: A modeling study. *J. Adv. in Modeling Earth Systems*, *11*(3), 712-731. doi: https://doi.org/10.1029/2018MS001519

Seifert, A., & Rasp, S. (2020). Potential and limitations of machine learning for modeling warm-rain cloud microphysical processes. *J. Adv. in Modeling Earth Systems*, *12*(12), e2020MS002301.

Shima, S.-i., Kusano, K., Kawano, A., Sugiyama, T., & Kawahara, S. (2009). The super-droplet method for the numerical simulation of clouds and precipitation: A particle-based and probabilistic microphysics model coupled with a non-hydrostatic model. *Quart. J. Roy. Met. Soc.*, *135*(642), 1307–1320.

Siewert, C., & Seifert, A. (2018). McSnow: Particle-based microphysics model in ICON. Poster at ICCARUS (ICON/COSMO/CLM/ART User Seminar), Offenbach, Germany. Retrieved from `https://download.dwd.de/pub/DWD/Forschung_und_Entwicklung/ICCARUS2018_presentations_PDF/Monday/P17M_Siewert.pdf` (last accessed 11/12/2023)

Ukkonen, P. (2022). Exploring pathways to more accurate machine learning emulation of atmospheric radiative transfer. *J. Adv. in Modeling Earth Systems*, *14*(4), e2021MS002875. doi: https://doi.org/10.1029/2021MS002875

Ukkonen, P., Pincus, R., Hogan, R. J., Pagh Nielsen, K., & Kaas, E. (2020). Accelerating radiation computations for dynamical models with targeted machine learning and code optimization. *J. Adv. in Modeling Earth Systems*, *12*(12), e2020MS002226. doi: https://doi.org/10.1029/2020MS002226

Ullrich, R., Hoose, C., Möhler, O., Niemand, M., Wagner, R., Höhler, K., … Leisner, T. (2017). A new ice nucleation active site parameterization for desert dust and soot. *J. Atmos. Sci.*, *74*(3), 699–717. doi: 10.1175/JAS-D-16-0074.1

Vobig, K., Stephan, K., Blahak, U., Khosravian, K., & Potthast, R. (2021). Targeted covariance inflation for 3d-volume radar reflectivity assimilation with the letkf. *Quart. J. Roy. Met. Soc.*, *147*(740), 3789–3805.

Walko, R. L., Cotton, W. R., Meyers, M., & Harrington, J. (1995). New rams cloud microphysics parameterization part i: the single-moment scheme. *Atmos. Res.*, *38*(1-4), 29–62.

Weisman, M. L., & Klemp, J. B. (1982). The dependence of numerically simulated convective storms on vertical wind shear and buoyancy. *Mon. Wea. Rev.*, *110*(6), 504–520.

Welss, J.-N., Siewert, C., & Seifert, A. (2023). Explicit habit-prediction in the Lagrangian super-particle ice microphysics model McSnow. *Authorea Preprints*. (submitted to J. Adv. in Modeling Earth Systems)

Xue, L., Fan, J., Lebo, Z. J., Wu, W., Morrison, H., Grabowski, W. W., … Rasmussen, R. M. (2017). Idealized simulations of a squall line from the mc3e field campaign applying three bin microphysics schemes: Dynamic and thermodynamic structure. *Mon. Wea. Rev.*, *145*(12), 4789 - 4812. Retrieved from `https://journals.ametsoc.org/view/journals/mwre/145/12/mwr-d-16-0385.1.xml` doi: 10.1175/MWR-D-16-0385.1

Zängl, G., Reinert, D., Rípodas, P., & Baldauf, M. (2015). The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quart. J. Roy. Met. Soc.*, *141*(687), 563–579.

Zeng, Y., Blahak, U., & Jerger, D. (2016). An efficient modular volume-scanning radar forward operator for NWP models: description and coupling to the COSMO model. *Quart. J. Roy. Met. Soc.*, *142*(701), 3234–3256.

Appendix C Open Research

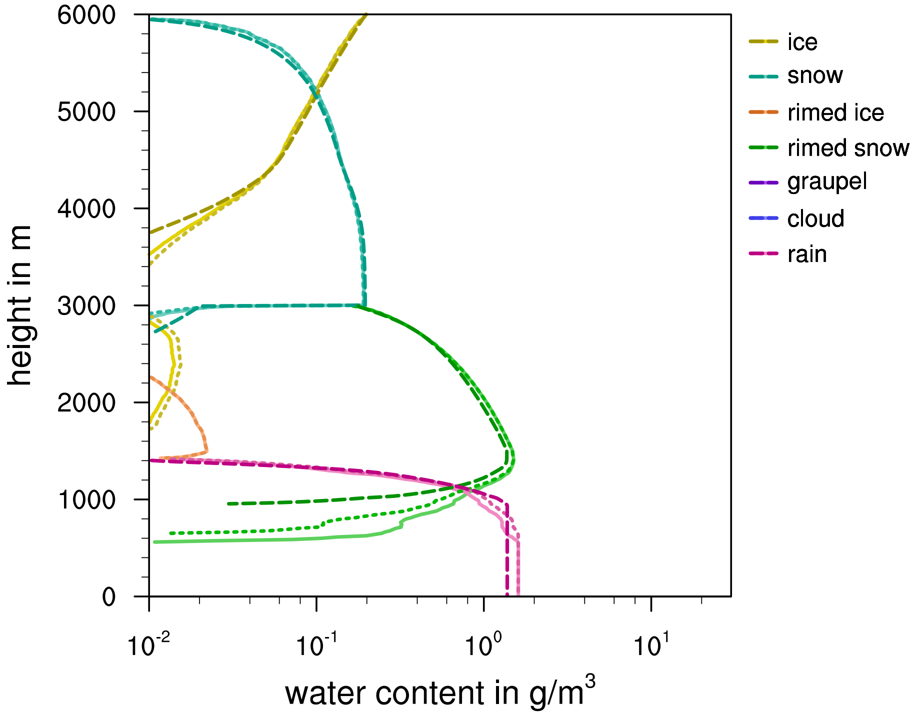
The training data, the Python notebooks and NCL scripts are provided at Zenodo `doi:10.5281/zenodo.10408950`. The Python notebooks are also publicly accessible at `https://gitlab.com/axelseifert/iceml`. The Zenodo archive does in addition include the Fortran modules containing all the newly developed ICON code for the ML-based P3-like microphysics scheme. The Lagrangian microphysics model McSnow is part of the ICON modeling framework, which is a joint effort of Deutscher Wetterdienst (DWD) and the Max Planck Institute for Meteorology (MPI-M). ICON licenses for scientific use are available at no cost at `https://code.mpimet.mpg.de/projects/iconpublic/`. Subsequently, the access to the McSnow GIT archive can be granted by A.S or C.S.

Acknowledgments

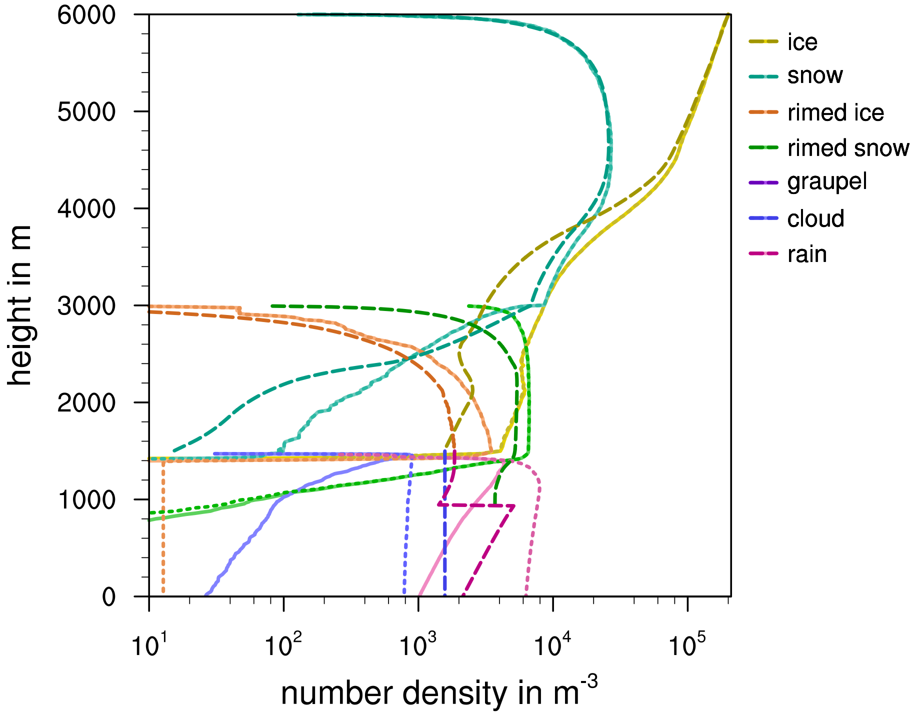
The authors thank Jens-Olaf Beismann for help with the implementation and vectorization of the neural nets in ICON and Leonhard Scheck for providing the Fortran code for the multilayer perceptron. The study contributes to the project FRAGILE of the SPP 2115 Fusion of Radar Polarimetry and Numerical Atmospheric Modelling Towards an

Improved Understanding of Cloud and Precipitation Processes funded by the Deutsche Forschungsgemeinschaft (DFG 492234709).

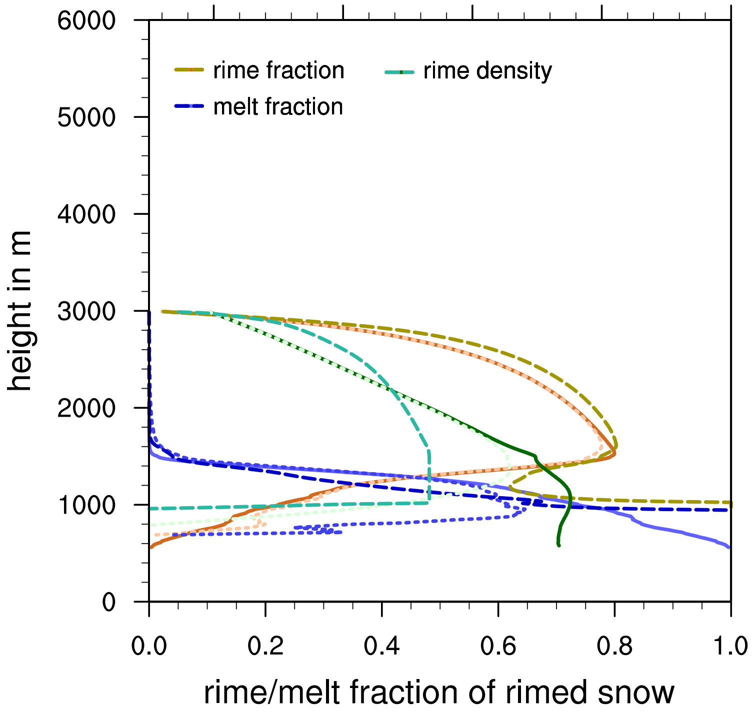
Figures 1-7.



6000
5000
4000
3000
2000
1000
0
height in m
10^{-2}
10^{-1}
10^{0}
10^{1}
water content in g/m^3
ice
snow
rimed ice
rimed snow
graupel
cloud
rain

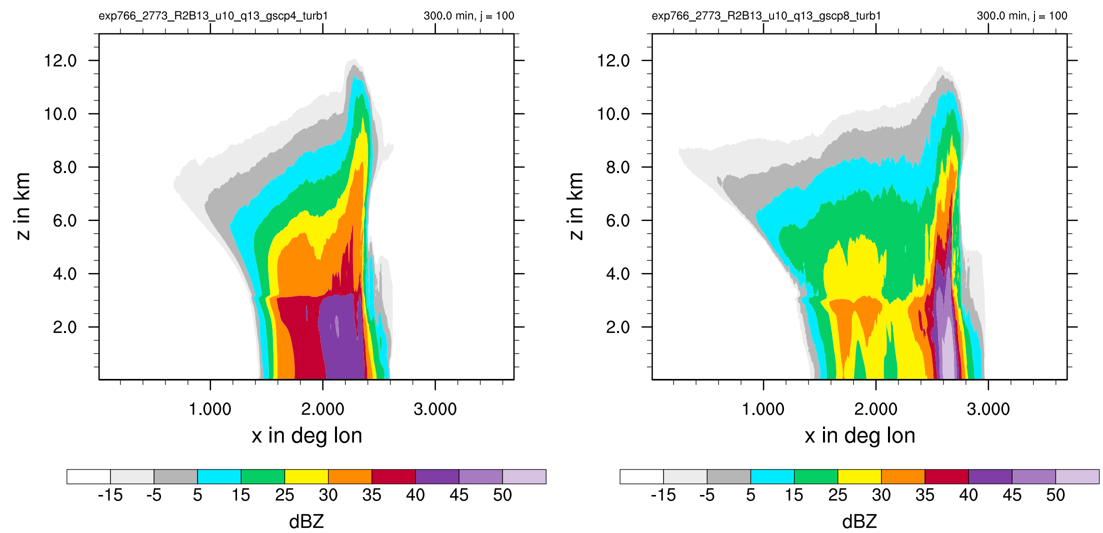


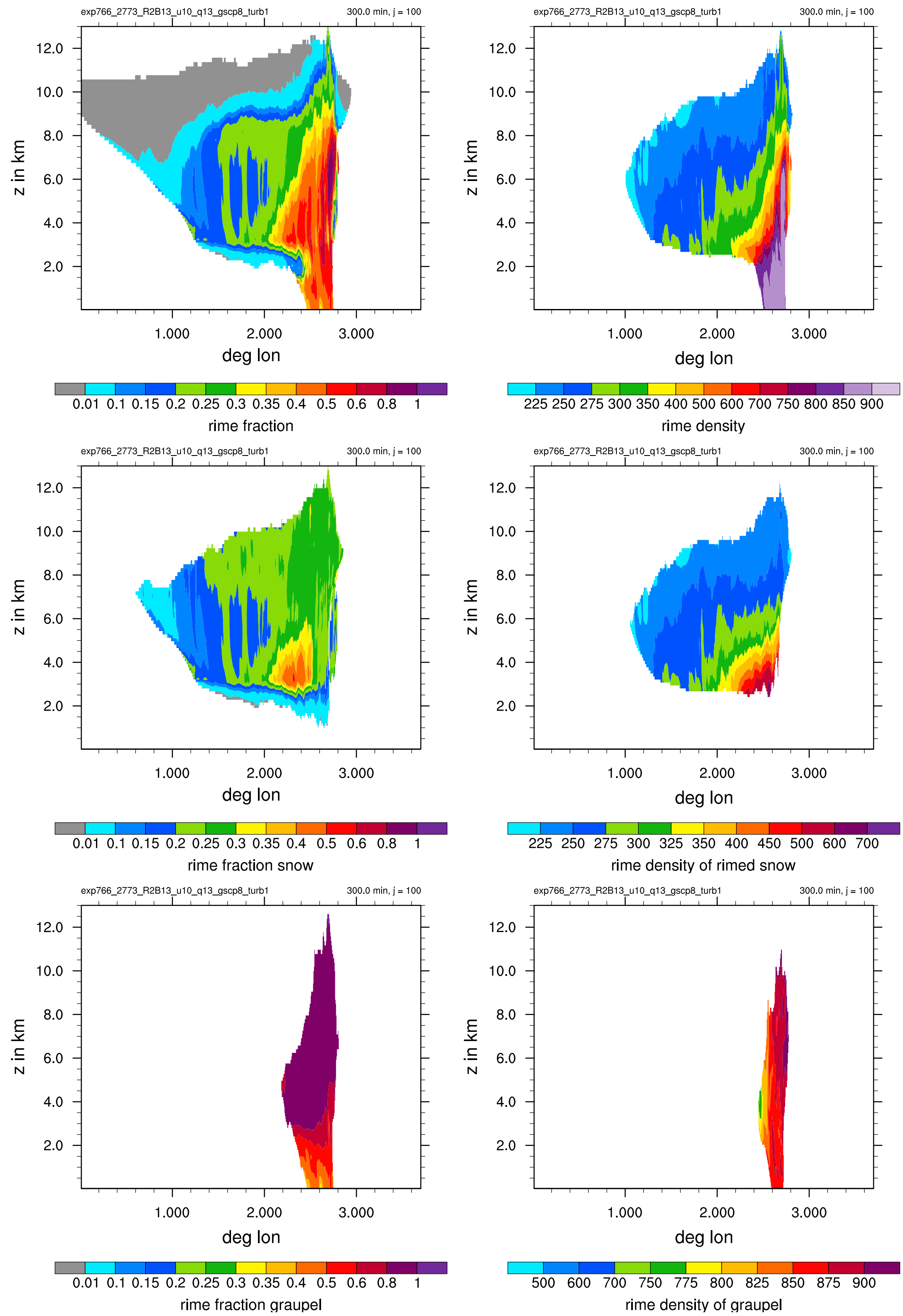
6000
5000
4000
3000
2000
1000
0
height in m
10^1
10^2
10^3
10^4
10^5
number density in m^{-3}
ice
snow
rimed ice
rimed snow
graupel
cloud
rain



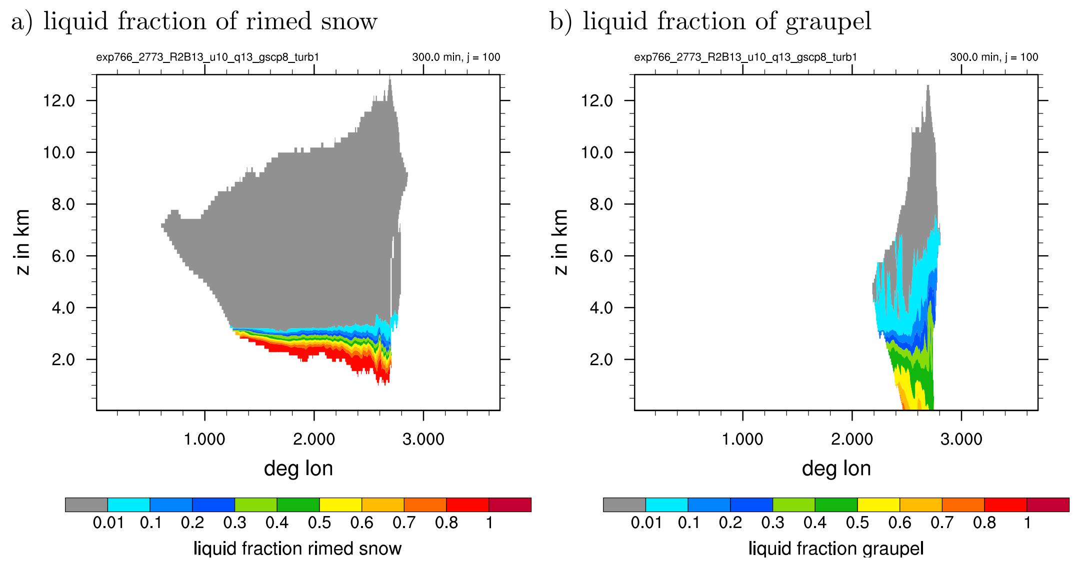
rime fraction
rime density
melt fraction
height in m
rime/melt fraction of rimed snow
6000
5000
4000
3000
2000
1000
0
0.0
0.2
0.4
0.6
0.8
1.0

a) SB two-moment b) ML-based P3-like two-moment

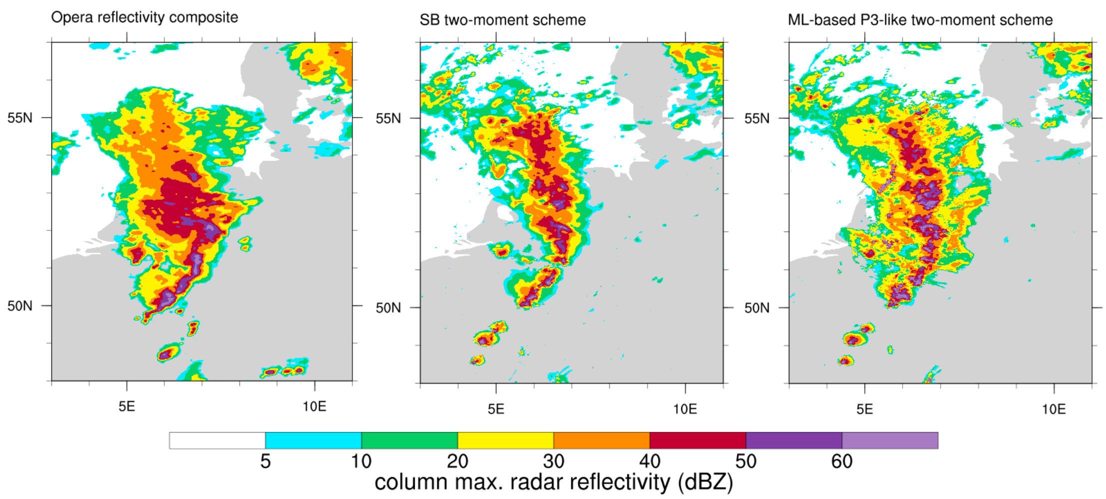




exp766_2773_R2B13_u10_q13_gscp8_turb1
300.0 min, j = 100
z in km
12.0
10.0
8.0
6.0
4.0
2.0
1.000
2.000
3.000
deg lon
0.01 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.5 0.6 0.8 1
rime fraction
225 250 275 300 350 400 500 600 700 750 800 850 900
rime density
rime fraction snow
225 250 275 300 325 350 400 450 500 600 700
rime density of rimed snow
rime fraction graupel
500 600 700 750 775 800 825 850 875 900
rime density of graupel



a) liquid fraction of rimed snow
exp766_2773_R2B13_u10_q13_gscp8_turb1
300.0 min, j = 100
z in km
12.0
10.0
8.0
6.0
4.0
2.0
1.000
2.000
3.000
deg lon
0.01 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 1
liquid fraction rimed snow
b) liquid fraction of graupel
exp766_2773_R2B13_u10_q13_gscp8_turb1
300.0 min, j = 100
z in km
12.0
10.0
8.0
6.0
4.0
2.0
1.000
2.000
3.000
deg lon
0.01 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 1
liquid fraction graupel


Opera reflectivity composite
SB two-moment scheme
ML-based P3-like two-moment scheme
55N
50N
5E
10E
5
10
20
30
40
50
60
column max. radar reflectivity (dBZ)

An ML-based P3-like multimodal two-moment ice microphysics in the ICON model

Axel Seifert[1], Christoph Siewert[1]

[1]Deutscher Wetterdienst, Offenbach, Germany

Key Points:

- Machine learning is successfully applied to build a complex bulk ice microphysics scheme by coarse-graining output of a Lagrangian particle microphysics model.
- The ML-based P3-like microphysics scheme improves the representation of the stratiform region of an idealized squall line compared to a classic two-moment scheme.
- The ML-based P3-like microphysics scheme runs stable and provides meaningful results in three-dimensional real-case simulations.

Corresponding author: Axel Seifert, `axel.seifert@dwd.de`

Abstract

Machine learning (ML) is used to build a bulk microphysical parameterization including ice processes. Simulations of the Lagrangian super-particle model McSnow are used as training data. The machine learning performs a coarse-graining of the particle-resolved microphysics to multi-category two-moment bulk equations. Besides mass and number, prognostic particle properties (P3) like melt water, rime mass, and rime volume are predicted by the ML-based bulk model. The ML-based scheme is tested with simulations of increasing complexity. As a box model, the ML-based bulk scheme can reproduce the simulations of McSnow quite accurately. In 3d idealized squall line simulations, the ML-based P3-like scheme provides a more realistic extended stratiform region when compared to the standard two-moment bulk scheme in ICON. In a realistic case study, the ML-based scheme runs stably, but can not significantly improve the results. This shows that machine learning can be used to coarse-grain super-particle simulations to a bulk scheme of arbitrary complexity.

Plain Language Summary

Numerical weather prediction and climate models need a description of unresolved cloud microphysical processes. Such microphysical parameterizations are usually formulated as systems of equations for bulk variables that describe the time evolution of clouds and precipitation. In this study, we use machine learning (ML) techniques to build such a parameterization. As input or training data simulations of a very detailed cloud model are used. This detailed model provides information not only on the mass and number of cloud particles but also other properties like the degree of melting or the mass of liquid drops frozen on the ice particles called rime mass. The machine learning approach can successfully construct the necessary statistical relations that are needed for microphysical parameterization. This parameterization is then tested in simulations of increasing complexity. The new ML-based scheme provides physically reasonable solutions and improves the simulation of a line of thunderstorms

1 Introduction

Developing parameterizations for numerical weather prediction (NWP) and climate models can be a tedious and time-consuming task (Jakob, 2010). Speeding up this development cycle is crucial for further progress in understanding and predicting regional climate change and improve NWP models to forecast hazardous and extreme weather events (Bauer, Stevens, & Hazeleger, 2021).

Machine learning methods hold the promise for a more rapid model development cycle, for example, through a semi-automatic workflow from highly-resolved reference simulations to coarse-grained and computationally efficient algorithms. Machine learning algorithms as an integral part of NWP and climate models may allow for better performance optimization and on-the-fly calibration with observations (Bauer, Dueben, et al., 2021). Machine learning methods have recently gained much attention in atmospheric modeling especially for emulators that help to improve the computational performance of the model (Ukkonen et al., 2020; Lagerquist et al., 2021; Chantry et al., 2021; Meyer, Grimmond, et al., 2022; Meyer, Hogan, et al., 2022; Ukkonen, 2022).

Here we take a rather straightforward approach to machine learning in that we use supervised learning with fully connected neural nets applied to individual physical process rates. This approach has the advantage that it is conceptually very similar to classic parameterizations, i.e., the result is an ODE system for the bulk variables (Seifert & Rasp, 2020; Gettelman et al., 2021). It also ensures mass conservation and allows a posteriori analysis of the ML representation of individual physical processes. In addition, it can be applied at different model time steps and even horizontal resolutions without having to re-train the ML model. The main disadvantage of this simple approach to ML is that it is not as computationally efficient as ML methods could be if applied in a more advanced and state-of-the-art framework, e.g., using a UNet++ architecture (Lagerquist et al., 2021) or recurrent neural nets (Ukkonen, 2022). Our ML approach is quite similar to the use of look-up tables for microphysical process rates. The use of look-up tables has a long tradition in cloud microphysical modeling (Walko et al., 1995; Feingold et al., 1998). To avoid the term look-up-table they are sometimes even called bin-emulating schemes in cloud modeling literature (Khain et al., 2015), which should not mask the fact these are still bulk schemes with their intrinsic limitations.

Many of the currently available bin and bulk microphysical schemes have problems in representing the stratiform region of mesoscale convective systems like squall lines (Morrison et al., 2009; Xue et al., 2017). This is often attributed to the fact that they are based on a limited number of particle types like snow and graupel, whereas in nature the transition from snow to graupel by riming, i.e. by accretion of supercooled cloud droplets, is continuous (Mosimann et al., 1994; Seifert et al., 2019). Morrison and Milbrandt (2015, MM15 hereafter) suggested to abandon such particle types completely and instead use prognostic particle properties (P3) especially the rime mass and the rime volume to represent riming. The additional rime volume is important to predict rime density. Instantaneous rime density is a function of temperature and Stokes number (Cober & List, 1993), but for a given particle the rime density depends on its history and therefore requires an additional prognostic variable. The importance of prognostic rime mass for the simulation of deep convection is also discussed in Aligo et al. (2018). In their original P3 scheme, MM15 abandoned the multimodal representation that comes with multiple particle classes, but later they presented a version of their scheme with multiple categories (Milbrandt & Morrison, 2016), and recently also a version of P3 with a triple-moment representation (Milbrandt et al., 2021). In addition, an extended variant of the P3 scheme with prognostic melt water on ice particles has been developed (Cholette et al., 2019, 2020, 2023). Hence, the P3 approach represents the state-of-the-art of bulk microphysical parameterizations for high-resolution NWP and climate models.

In the following, we explore whether we can derive or 'learn' a P3-like scheme from Lagrangian super-particle simulations using standard machine learning methods. The aim is to build a semi-automatic workflow that generates a bulk microphysical scheme based on some a priori choices and simulations of the super-particle model McSnow.

The paper is organized as follows: In section 2 we introduce the basic assumptions of the new ML-based P3-like microphysics scheme. In section 3 the super-particle model McSnow and the simulations that serve as training data are described. The actual structure of the training data and the training process are discussed in section 4. Section 5 presents a comparison of McSnow and the ML-based bulk model. In section 6 the ML-based model is applied to idealized three-dimensional squall line simulations with the ICON model. Section 7 presents a realistic case study with ICON and a comparison of the ML-based bulk models with a classical two-moment bulk microphysics scheme. The paper ends with a Summary and Conclusions.

2 A priori choices for the ML-based P3-like microphysics

To build an ML-based microphysics scheme, we have to make some a priori choices regarding the number of hydrometeor categories and the corresponding prognostic bulk variables. In contrast to MM15, we think that multimodality is ubiquitous in clouds because the various pathways for the formation of precipitation-sized particles lead to the co-existence of various modes or particle types. Hence, our scheme has multiple particle categories, but they have a clear and physically-based definition in terms of their formation mechanism. The cloud ice category comprises primary ice particles (monomers) which have only grown by depositional growth. Unrimed snow are aggregates of these primary crystals. Those two categories have only a two-moment representation with no additional properties. Then we have three categories that carry rime mass and rime volume: rimed ice, rimed snow, and graupel. Whereas rimed ice and rimed snow are simply the rimed monomers and rimed aggregates, the graupel originates from freezing of raindrops.

The latter two categories have a prognostic liquid water mass to explicitly represent melting and wet growth. Carrying unrimed ice (snow) and rimed ice (rimed snow) separately may sound unnecessary, given that we have prognostic rime mass, but due to the patchiness of supercooled liquid water a co-existence of rimed and unrimed particles in the same grid volume is not impossible. The hydrometeor categories and the corresponding prognostic variables of the ML-based P3-like scheme are summarized in Table 2. The scheme has overall 23 prognostic variables: 18 for the ice phase, 4 for the liquid phase, and one additional tracer for tracking activated ice nuclei (Köhler & Seifert, 2015). Note that the bulk classification in McSnow is different from the classes of the ML-based P3-like scheme in that the McSnow classification would allow a conversion from snow to graupel. In fact, this process is contained in the training data, but for the ML-based P3-like bulk scheme described in the current study, we decided not to allow snow-to-graupel conversion. Note that the particle classification in McSnow is only a diagnostic to analyze the simulations and generate training data for a bulk model. The bulk classification does not affect the microphysical processes in McSnow, which is by construction continuous and class-free.

With 23 prognostic variables and a high level of complexity, this ML-based scheme is not primarily aimed at operational NWP, where computational efficiency is of the essence

Table 1. Prognostic variables of the super-particle model McSnow and corresponding bulk variables of the ML-based P3-like scheme. Here ρ_i is the material density of ice and χ is the super-particle multiplicity divided by air volume. The hydrometeor categories are defined in Table 2.

Prognostic variables of McSnow super-particles		
variable	symbol	note
ice mass	m_i	increases by depositional growth
rime mass	m_r	increases by riming
rime volume	V_r	changes obey instantaneous rime density
liquid mass	m_ℓ	increases by melting or collection of liquid drops
frozen mass	m_f	increases by freezing of liquid mass
monomer number	N	number of ice monomers
multiplicity	χ	number of real particles per super-particle
Prognostic variables of ML-based P3-like bulk scheme for a hydrometeor category k		
mass density	q_k	sum of $(m_i + m_r + m_f + m_\ell)\chi$
number density	n_k	sum of χ
rime mass	ψ_k	sum of $(m_r + m_f)\chi$
rime volume	ϕ_k	sum of $(V_r + m_f/\rho_i)\chi$
liquid mass	ℓ_k	sum of $m_\ell\chi$

but in cloud modeling, regional climate research and other applications that may care about a good representation of cloud microphysics. The large number of prognostic variables make this scheme rather complicated, but quite suitable as a test case and proof-of-concept for the ML approach to parameterization development.

3 Super-particle simulations

The Lagrangian super-particle model McSnow (Brdar & Seifert, 2018) makes use of the Monte-Carlo algorithm of Shima et al. (2009) to simulate the collision and aggregation processes of hydrometeors. The super-particle approach allows for a direct representation of the evolution of the properties of individual hydrometeors. To do so, McSnow carries multiple variables to describe each hydrometeor. These are the hydrometeor state variables ice mass, rime mass, rime volume, number of monomers, frozen mass, and liquid mass (see Table 1). The ice mass increases due to depositional growth and determines the maximum dimension of the particle with the help of an empirical m-D relationship. Hence, in this configuration of McSnow, we do not employ the habit pre-

Table 2. Overview of the ML-based P3-like two-moment bulk microphysics scheme. The degree of riming ξ is here defined as $\xi = (m_r + m_f)/(m_i + m_r + m_f + m_\ell)$ including the frozen mass m_f.

class	variables	McSnow classification
unrimed ice	q_i, n_i	$N = 1$ and $\xi = 0$
unrimed snow	q_s, n_s	$N > 1$ and $\xi = 0$
rimed ice	q_{ri}, n_{ri}, ψ_{ri}, ϕ_{ri}	$N = 1$ and $0 < \xi \leq 0.95$
rimed snow	q_{rs}, n_{rs}, ψ_{rs}, ϕ_{rs}, ℓ_{rs}	$N > 1$ and $0 < \xi \leq 0.95$
graupel	q_g, n_g, ψ_g, ϕ_g, ℓ_g	$\xi > 0.95$
cloud droplets	q_c, n_c	$m_i + m_r + m_f = 0$ and $r < 40\ \mu$m
raindrops	q_r, n_r	$m_i + m_r + m_f = 0$ and $r \geq 40\ \mu$m

diction of Welss et al. (2023). Rime mass and rime volume increase due to collision with supercooled liquid drops. The instantaneous rime density is parameterized following Cober and List (1993). The number of monomers increases by aggregation, i.e., collection of other ice particles. Finally, melting in McSnow is based on Rasmussen et al. (1984a, 1984b) and Rasmussen and Heymsfield (1987b, 1987a). Melting increases the liquid mass m_ℓ, which is a prognostic variable for each individual super-particle. Freezing of liquid drops uses a probabilistic interpretation of the parameterization of Barklie and Gokhale (1959). As secondary ice production, only Hallet-Mossop rime splintering is currently considered in McSnow (Hallett & Mossop, 1974; Field et al., 2017).

We assume that all hydrometeors fall with their terminal fall velocity v_t. The terminal fall velocities of all hydrometeors and also the collision efficiency E_c of all possible mutual collisions are parameterized using the approach of Böhm (1992a, 1992b, 1992c, 1994, 1999, 2004). Using Böhm's theory provides a continuous and physically consistent dependency of the hydrometeor properties, like v_t or E_c, and consequently the collision kernel K on the hydrometeor state variables. Welss et al. (2023) provide a more detailed discussion of Böhm's theory in the framework of McSnow. Special considerations are necessary for the sticking efficiency of unrimed and rimed snow and graupel. Usually, the sticking efficiency of snow is parameterized as a function of temperature whereas the sticking efficiency of graupel is most often assumed to be small and constant. In McSnow a continuous parameterization as a function of temperature and degree of riming is applied, which is specified in Appendix A.

McSnow can be used in a one-dimensional version as in Brdar and Seifert (2018) and Bringi et al. (2020) and in two- and three-dimensional simulation as part of the ICON model (Siewert & Seifert, 2018). As training data for the machine learning approach, we need a broad range of environmental and microphysical parameters. Hence, two- or three-dimensional simulations would be far too expensive. Even the one-dimensional model is inefficient because it needs several hours of simulation time to spin up a quasi-equilibrium state. To overcome these obstacles, we have implemented a simple zero-dimensional box model that approximates the quasi-equilibrium state of the one-dimensional McSnow, but is computationally cheaper. The zero-dimensional McSnow describes a population of hydrometeors initialized as pristine ice of unrimed monomers that fall through a prescribed atmosphere. The sedimentation velocity of the box model is equal to the mass-weighted terminal fall velocity of all hydrometeors in the box. The atmospheric profile is the same as in Brdar and Seifert (2018), their Figure 6. While the box is falling through the atmosphere, the hydrometeors grow by depositional growth and mutual binary collisions. They encounter a layer of supercooled liquid drops and grow by riming. Without the presence of liquid water, ice particles start melting when they reach the 0°C level. As melting in McSnow is formulated by a quasi-equilibrium energy budget, large ice particles can reach the wet growth regime in regions of high liquid water content. In the wet growth regime, a liquid water layer exists on the ice particles even at temperatures below 0 °C.

The simulations described in the previous paragraph mimic the microphysical processes in a stratiform cloud including the stratiform regions of convective systems. In convective updrafts other processes, like freezing of raindrops and riming with raindrops are important or even dominant, which are not well represented in those simulations. To sample the microphysical processes as they occur in convective clouds, the same atmospheric profile is used, but the box model is initialized near the surface with an arbitrarily chosen upward velocity of 5 m/s. This leads to the formation of raindrops in the parcel, which subsequently freeze and start riming. These simulations provide the data for microphysical processes as they happen within updraft cores of convective systems. When the updraft parcel reaches a height of $0.95\,h_{\mathrm{top}}$, the updraft ends, and the parcel enters the regular sedimentation mode described above, where it falls with the mass-weighted sedimentation velocity of the hydrometeors. This is necessary to provide, for example, training data for the melting of graupel.

Table 3. Parameter hypercube of McSnow simulations for the training data. The random sampling is based on uniform random variables $r \in [0.1]$. The parameters for additive sampling are $c_h = 1$ km and $c_r = 5$ μm. Not all heights h_1 and Δh_2 have been used with all domain tops h_{top}.

	Basic McSnow simulations		
variable	range of values	random sampling	unit
ice supersaturation S_i	0.0, 0.2, 0.5	$S_i^* = S_i\,(1+r)$	-
height h_1	500, 1000 1500	$h_1^* = h_1 + C\,r$	m
height Δh_2	500, 1500, 2000	$h_2^* = h_1 + \Delta h_2\,(1+r)$	m
droplet radius r_c	5, 10, 15, 20, 25, 30	$r_c^* = r_c + Cr$	μm
ice number density N_i	10, 20, 40, 80, 160, 320, 640	$N_i^* = N_i(1+r)$	dm^{-3}
ice water content Q_i	0.1, 0.2, 0.4	$Q_i^* = Q_i(1+r)$	g cm^{-3}
cloud water content Q_c	0.1, 0.2, 0.4, 0.8	$Q_c^* = Q_c(1+r)$	g cm^{-3}
domain top h_{top}	5500, 6500, 7500, 8500, 9500	-	m
	Updraft McSnow simulations		
variable	range of values	random sampling	C
ice supersaturation S_i	0.0, 0.2	$S_i^* = S_i\,(1+r)$	-
height h_1	500, 1500, 2500	$h_1^* = h_1 + C\,r$	m
height Δh_2	2000, 4000	$h_2^* = h_1 + \Delta h_2\,(1+r)$	m
droplet radius r_c	5, 10, 15, 20, 25, 30	$r_c^* = r_c + Cr$	μm
ice number density N_i	10, 20, 40, 80, 160, 320, 640	$N_i^* = N_i(1+r)$	dm^{-3}
ice water content Q_i	0.1, 0.2, 0.4	$Q_i^* = Q_i(1+r)$	g cm^{-3}
cloud water content Q_c	0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4	$Q_c^* = Q_c(1+r)$	g cm^{-3}
domain top h_{top}	6000, 7000, 9000	-	m

4 Training data and ML results

To build a bulk microphysics scheme using a standard machine learning workflow, we first choose the prognostic variables of the desired scheme. Here we decided on a two-moment approach with seven particle categories, and rimed ice, rimed snow and graupel have additional prognostic variables for rime mass, rime volume, and liquid mass (see Table 2). As shown by Seifert and Rasp (2020, SR20 hereafter) the ML approach has no advantages for the warm-rain processes that determine the growth of cloud droplets and raindrops. Hence, we use existing parameterizations for the warm-rain processes based on Seifert and Beheng (2001) and Seifert (2008). Ice nucleation is parameterized using semi-empirical approaches and is not modified by machine learning. The ice nucleation active site (INAS) density approach of Ullrich et al. (2017) is used for heterogeneous ice nucleation, whereas homogeneous ice nucleation follows Kärcher et al. (2006).

This leaves 55 process rates that need to be parameterized by the machine learning approach. Those process rates include depositional growth and melting of ice particles, and all collisional processes like aggregation among ice particles and riming of ice by collection of cloud droplets or raindrops. All those processes depend, on one hand, on the physical properties of the hydrometeors (terminal fall velocity, particle size distribution) and change, on the other hand, the bulk properties of the ice categories. Tables 4-6 summarize all those processes, the input variables (predictors or features) of the neural net, and the predicted process rates (output or labels). In addition to the 55 process rates (networks 1-27), the bulk sedimentation velocities are needed to quantify the precipitation fluxes of the particle categories (networks 28-32). To compare with observations, an estimate of the radar reflectivity is needed that depends primarily on the 2nd mass moment. Hence, this also has to be estimated by a neural net (networks 33-37). Finally, for a consistent coupling with radiation, we need effective radii for each category (networks 38-42). The diagnostic neural nets are necessary because neither McSnow nor the ML-based model make any a priori assumptions about particle size distributions. Hence, all bulk particle properties have to be learned from the training data.

All these networks are simple dense fully-connected multilayer perceptrons and the size for each network is specified in the Tables. The size has been determined by hyperparameter study and subsequent testing with the 1d and 2d simulations that are described in the following sections. The network size is relevant for the computation time needed and large neural nets would make the scheme considerably more expensive. This is especially important for processes that occur almost everywhere in a cloud, like depositional growth, and less so for very special processes that occur rarely like melting of graupel.

All 42 neural nets are regression models and directly provide the required physical variable (process rate, sedimentation velocity, etc.). This proved to work well in this case for all processes except for the self-collection of unrimed snow. Specifically for this process, we found that a two-step approach as described by Gettelman et al. (2021) is indeed beneficial and improves the performance of the overall scheme. The two-step approach uses a classifier network, which first determines whether the process rate is non-zero, followed by a regression network that estimates the actual process rate. Network no. 43 estimates the probability of self-collection of unrimed snow. Only where this probability exceeds 0.5 the regression network no. 7 is applied to calculate the corresponding process rates.

For all 55 process rates and the additional variables like sedimentation velocities, radar reflectivities, and effective radii, training data needs to be calculated from the McSnow output. This is straightforward and involves only sums over super-particles. Hence, no additional assumptions, choices, or approximations are necessary at this stage. For a derivation, we refer to section 3 of SR20, who detail this step for the warm-rain processes.

To generate a broad range of training samples, McSnow simulations are done for various atmospheric and microphysical conditions. Hence, we change the assumptions in the atmospheric profiles like domain height, height and thickness of the liquid layer, initial ice water content and ice number density of the parcel, the liquid water content and mean radius of the cloud droplets in the liquid layer, and the ice supersaturation outside of the liquid layer. Details are given in Table 3. This constitutes a multidimensional gridded hypercube from which we draw random samples. We did not perform a Latin hypercube sampling but drew samples in each box, which could be described as full hypercube sampling. Overall we have performed more than 20000 McSnow simulations for this study. The number of training and testing/validation samples for each process is given in Tables 4-6. These are of order 10^6 training samples for common processes like depositional growth, but the number drops down to 10^5 for many riming processes, and some processes that are rare or occur only in thin layers like self-collection of rimed ice or ice multiplication have only a few thousand training samples.

For the training of the individual neural nets Tensorflow 2.1 has been used. We randomly select 70 % of the data for training, 15 % for testing during the training process, and 15 % for validation after the training process. This split of the data is done for each process independently. We choose the mean squared error (MSE) as loss function, ReLU activation, an initial learning rate of 1e-3, the Adam optimizer, and early stopping (with a patience parameter between 5 and 10 depending on the process, and restoration of the best weights). All those choices are fairly standard for such simple regression networks with Tensorflow, but they seem to work well in our case. The training, validation, and testing data are standardized using the mean and standard deviation to ensure that all features have zero mean and a standard deviation of one. Hence, we apply the transformation $\check{\zeta} = (\zeta - \bar{\zeta})/\sigma_\zeta$, where ζ is a feature vector with mean $\bar{\zeta}$ and standard deviation σ_ζ, and $\check{\zeta}$ is the standardized value of the feature.

For most process rates, especially the collision rates, all dependencies are learned from the data. For some thermodynamic processes, we decided not to learn linear dependencies that are well-known. For example, the deposition rate is a linear function of the supersaturation S_i. We do not have to infer that from data. Hence, this linear dependency is removed from the training data. This explains why NNs 1-5 have no dependency on water vapor or supersaturation. Similarly, for melting rates of the internal melting, i.e. the melting within one particle category that converts ice or rime mass to liquid mass, we assume a linear dependency on the temperature deviation from the melting point. The latter is an approximation but works better in this case, because the melting layers can be thin and are then not sufficiently sampled by the training data.

For most of the neural nets the mean absolute error (MAE) and mean squared error (MSE) become sufficiently small. A value below 0.1 for MAE and below 0.05 for MSE is already sufficient given the complexity of the problem and the uncertainty of the assumptions made in such schemes. For some processes that are complicated but lack training data, the errors remain larger. Some processes involving rimed ice show the largest uncertainty because rimed ice exists only in rather thin layers which limits the number of training samples. At the same time, rimed ice can change its properties quite strongly due to the variability in rime density. At least, for the most frequent and most important processes it is possible to have enough training data to achieve a good and robust approximation.

5 Comparison with super-particle reference

A first and already quite challenging test for the ML-based model is to reproduce the training data when the individual process rates are combined into a system of ordinary differential equations (ODEs). This means that the process rates trained independently in the previous section constitute an ODE system and should reproduce the bulk variables of the McSnow simulation when used in concert. It is far from trivial that this works, because it requires that the process rates are sufficiently well approximated in a large part of the phase space. Hence, although we use the training data or very similar simulations for validation, this is a meaningful test for the ML-based model. The four-equation warm-rain model investigated in SR20 partly failed this test, because although the results of the ML model were reasonable approximations of the reference model, they were inferior to a well-established analytic parameterization.

Figures 1 and 2 present the results of such an ODE test of the ML-based microphysics. Shown are vertical profiles of water content and the number density of the various hydrometeor categories. As explained earlier, the vertical profiles are equivalent to time series, because the parcel (box model) falls through a prescribed atmospheric profile. The parcel starts with only unrimed cloud ice as initial condition. Soon unrimed snow forms by aggregation and starts to dominate the ice water content at 5000 m height, at 3500 m almost all cloud ice has been depleted. The unrimed snow reaches the liquid water layer at 3000 m height and almost instantaneously becomes rimed snow. Small amounts of unrimed ice and rimed ice exist within the liquid water zone. Unrimed ice can exist there as long the particles are smaller than the riming onset of approximately 100 μm. At 1500 m the particles reach the 0 °C level and the rimed snow melts into raindrops. The ML-based model can reproduce this archetypical behavior very well and the match with the original McSnow data is very good. The only small error is that the ML-based model is not able to produce sufficient amounts of unrimed and rimed ice in the liquid water zone. Note that the dotted lines, which represent the ODE system with the bulk process rates that would serve as training data (if this simulation would actually be part of the training data, which it is not), match the McSnow output even better and do capture the unrimed and rimed ice in the liquid layer. Hence, this information is in principle contained in the training data. The profiles of the number densities are much more complicated and show larger differences between McSnow and the ML-based bulk model. Nevertheless, the qualitative behavior is captured well by the ML-based model. The number densities show that the ML-based model does have some unrimed and rimed ice in the liquid water zone, but it is a factor 2-3 too low compared to McSnow. The profiles of the number densities also reveal some approximations that we made in the formulation of the bulk process rates at and below the melting level, and, hence, neither the training data (dotted lines) nor the ML-based model (dashed lines) match the reference of McSnow (solid lines) perfectly for melting particles and raindrops. Figure 3 shows some of the bulk particle properties of rimed snow for the same simulation as Fig. 1. Here the rime fraction matches quite well between McSnow and the ML-based bulk scheme, but the rime density of rimed snow shows initially a too rapid increase but then it flattens off and does not reach the same values as the reference. Melting happens in a rather thin layer of only 500 m, but at least the ML-based model does show a reasonable increase in melt fraction of rimed snow and, hence, captures the thickness of the melting layer quite well. Overall, the ML-based is able to pass this first test, because it provides a reasonable evolution of the microphysical variables including the prognostic particle properties like rime fraction and melt fraction for this archetypical but highly idealized case.

6 Squall line simulation with ICON

To perform idealized squall line simulations with the new ML-based P3-like microphysics, the scheme has been implemented in the ICON model (Zängl et al., 2015). For the ML-based scheme, the neural nets need to be evaluated in ICON as part of the model physics. To achieve this, the coefficients of all neural nets are stored in NetCDF files, which can easily be read into ICON. The evaluation of the neural nets, often called inference, is done using Fortran code originally developed for an ML-based satellite forward operator (Scheck, 2021). Other possible coupling strategies for using machine learning in ICON are for example discussed in Arnold et al. (2023).

To improve the efficiency of the implementation, especially on the NEC Aurora vector architecture currently in use at DWD, index lists are generated for each process. The index lists collect the grid points at which the input variables relevant for that process are non-zero, and the neural nets are then only evaluated where a non-zero process rate can be expected. This improves the efficiency not only on vector machines because many processes are non-zero only in very small parts of the three-dimensional domain, e.g., in deep convective updrafts with supercooled liquid water in case of riming processes. With this implementation the computational effort is bearable, but the scheme is considerably slower than the SB2006 two-moment microphysics. To some extent, because it has more prognostic variables (23 compared to 13), but the most expensive part of the ML-based scheme is in fact the inference of the neural nets. Some more implementation details are given in Appendix B.

To simulate a 3D idealized squall line the sounding of Weisman and Klemp (1982) is used with a linear wind profile from the surface to 2500 height and constant wind speed of 10 m/s above similar to Rotunno et al. (1988). The water vapor mixing ratio near the surface is 13 g/kg. The ICON simulation applies an R2B13 triangular icosahedral grid corresponding to an equivalent grid spacing of 308 m and a limited-area domain of 1.5 degree $\times$ 6.0 degree in the horizontal. The vertical grid has 128 levels with a domain top at 23 km and a damping layer starting at 20 km height. The TKE-based Mellor-Yamada level 2.5 boundary layer scheme is applied for vertical diffusion in combination with a 2D Smagorinsky closure in the horizontal. The Phillips et al. (2008) ice nucleation parameterization is used with constant number densities for dust, soot and organics given as $n_{\mathrm{dust}} = 1.6 \times 10^6$ m^{-3}, $n_{\mathrm{soot}} = 25 \times 10^6$ m^{-3} and $n_{\mathrm{orga}} = 30 \times 10^6$ m^{-3} similar to the 'high IN' setting of Seifert et al. (2012). The CCN activation is parameterized based on Segal and Khain (2006) with $N_{\mathrm{CN}} = 500 \times 10^6$ m^{-3}. In the following ICON simulations with the bulk two-moment scheme of Seifert and Beheng (2006, SB hereafter) are compared with the new ML-based P3-like bulk microphysics schemes.

The spatial structure of the squall line can be quantified with help of the radar reflectivity factor (dBZ). Observations often show a bimodal structure with high dBZ values in the convective core and a secondary weaker maximum in the trailing stratiform regions (see e.g. Figure 3 of Xue et al. (2017)). The separation of these two regions with a clear minimum in between, is difficult to capture with atmospheric models as discussed by Morrison et al. (2009) and Xue et al. (2017). Figure 4 shows vertical cross-sections of the radar reflectivity factor for ICON simulations with the SB scheme and the ML-based P3-like scheme. Using the SB scheme results in a relatively narrow squall line, which is dominated by the convective core and has no clear separation in convective and stratiform region. This is different for the ML-based P3-like scheme, which supports a more extended stratiform region with a more pronounced secondary maximum. Both microphysical schemes provide a reasonable squall line structure, but the ML-based scheme can alleviate some of the deficiencies of the SB schemes.

To achieve this improved spatial structure the ML-based scheme needs to be able to predict the evolution of the physical properties of the hydrometeors in the squall line. That the ML-based scheme is able to do this, is shown in Figure 5. The bulk rime fraction shows high values within the convective core where heavy riming occurs, the rime fraction decreases continuously within the stratiform region. This is reasonable, because little riming should happen outside the convective core, and particles with higher rime fraction have higher fall velocity and fall out more quickly. Decomposing the rime fraction in snow and graupel categories reveals that the rime fraction within the convective core is dominated by graupel, which has a rime fraction larger than 0.8. The stratiform region is almost only rimed snow and the rime fraction of rimed snow shows a maximum just behind the convective core. From there it decreases continuously because strongly rimed particles are removed by sedimentation. A similar structure is seen in rime density and the explanation is similar in the sense that riming is happening in the convective core and particles with high rime density have higher fall speeds and are removed by sedimentation. This suggests that the ML-based P3-like scheme does in fact capture the main physical processes and dependencies correctly. A more detailed analysis would require validation with in-situ observations or polarimetric radar data, which is beyond the scope of the current study.

Another interesting feature of the ML-based P3-like scheme is the explicit liquid mass of the rimed particle categories. For the squall line case, the liquid water fraction of rimed snow and graupel is shown in Figure 6. For rimed snow, the melting layer is roughly 2 km deep near the convective core and becomes thinner in the stratiform regions. This is again easily understood due to the larger and more heavily rimed particles closer to the convective core. Graupel reaches the ground in the convective core with a liquid fraction of 0.5. For graupel, the ML-based model predicts wet graupel up to 6 km height. This is physically possible because in zones of high supercooled liquid water, the riming rate becomes so large that the latent heat of freezing can no longer be dissipated by diffusion. This regime is called wet growth and is usually not represented in bulk microphysics schemes. The marginal liquid ratios between 0.1 and 0.01 below 8 km height are due to pockets of supercooled liquid water that can occur locally. Based on the results above, we can conclude that the ML-based P3-like scheme passed this idealized squall line test and delivers the improvement that can be expected from the P3 approach.

7 Mesoscale simulation with ICON

Machine learning models that are trained on simulation data may work well in idealized simulations as the squall line of the previous section, but can nevertheless fail when applied in a real-world situation. Hence, the next test is an example of an actual numerical weather prediction case using the ICON-D2 configuration of Deutscher Wetterdienst (DWD) similar to the operational regional forecast. The operational NWP system at DWD consists of a global ICON model, currently at 13 km grid spacing, with a European two-way nest at 6.5 km grid spacing (called ICON-EU), and the regional ICON-

D2 with approximately 2 km grid spacing over central Europe (Reinert et al., 2023). ICON uses an icosahedral unstructured mesh, which for ICON-D2 has 542040 cells on each of the 65 model levels, and a vertically stretched grid. For ICON-D2 the vertical grid spacing in the lowest levels is smaller than 100 m, but near the tropopause, it is approximately 500 m (Reinert et al., 2023). The domain top is at 22 km height. Operationally ICON-D2 still uses a one-moment microphysics, but the pre-operational rapid update cycle (RUC) applies the SB two-moment microphysics. The RUC spins off from the one-moment analysis at 0 UTC and performs its own analysis with the two-moment scheme using a local ensemble transform Kalman filter (LETKF, Schraff et al. (2016); Vobig et al. (2021)). Here we use the RUC analysis from 12 UTC to initialize forecasts for the afternoon of 19 May 2022. Note that the analysis has only been done with the SB two-moment scheme, not with the new ML-based P3-like scheme. The latter is currently not possible, because it would require coupling the ML-based scheme with the radar forward operator EMVORADO (Zeng et al., 2016), which is beyond the scope of the current study. Forecasts are performed for 6 hours and compared with the European Opera radar reflectivity composite. We use a radar reflectivity factor in Rayleigh approximation for both microphysics schemes to allow a fair comparison. For the ML-based microphysics scheme the neural nets 33-37 provide the second mass moment of the particle size distribution, which is required for the radar reflectivity factor in Rayleigh approximation.

Figure 7 presents the column maximum radar reflectivity for 13:30 UTC. Note that the ICON-D2 domain used for these simulations is considerably larger than the area shown in the Figure. The Opera composite shows a squall line over the Netherlands and Belgium approaching Germany. The southern end of the line shows a narrow convective region with high reflectivity values, to the north a larger stratiform region is visible. The ICON-D2 forecast with the SB two-moment scheme captures the overall structure of the convective system, but the convective cores are too weak and the stratiform region is too narrow and not as extended as in the observations. These are typical biases of ICON-D2 with the SB two-moment scheme, which are quite pronounced in this case. With the ML-based P3-like scheme the convective line at the southern end of the convective complex is even weaker, although higher reflectivity values occur within active convective cores. The stratiform region is more extended compared to SB but is more symmetric around the convective line and does not resemble the observations better than the simulation using SB microphysics. Hence, in contrast to the idealized squall line, the ML-based P3-like scheme does not improve over the SB scheme in this real-case application. The improved structure in the idealized squall line simulation is only apparent at the high spatial resolution of the 308 m mesh and deteriorates on coarser grids (not shown). Increasing the resolution of the real-case simulation to a sub-km mesh would unfortunately be too costly with the ML-based microphysics as it is currently implemented. Hence, the result of this case study is undecisive. The ML-based P3-like scheme is stable and provides a reasonable representation of the mesoscale convective system, but it can not improve over the SB two-moment scheme in this case. A simple explanation for this result is that the microphysics scheme is not the limiting factor for the forecast quality in this case. It is very likely that the model dynamics and especially the boundary layer scheme play an important role and contribute to the deficiencies of this ICON forecast.

8 Summary and Conclusions

Machine learning has been applied to build a complex bulk microphysics scheme, which predicts not only particle mass and number but detailed physical properties like rime mass, rime volume, and liquid mass following the P3 approach of Morrison and Milbrandt (2015). Training data has been generated using idealized simulations with the super-particle model McSnow. Hence, the machine learning performs a coarse-graining of the detailed McSnow data to a bulk microphysics scheme. The human role in this process is twofold: First, to make an a priori choice of the prognostic equations, i.e. the number of particle categories and the prognostic variables for each category. Second, to design the McSnow simulations that provide the training data. Based on these two preparatory steps, the machine learning workflow is almost automatic and does not require much human intervention, except for some limited hyperparameter tuning. Standard regression neural nets are sufficient for most processes. Only for the self-collection of unrimed snow, we found that a two-step classifier-regression approach as recommended by Gettelman et al. (2021) is superior to using only a regression neural net. The ML-based P3-like microphysics scheme has been implemented in the ICON weather and climate model using Fortran code for the inference of fully connected neural nets.

The ML-based P3-like microphysics scheme has passed three relevant tests: First, it can reproduce simulations similar to the training data, which requires that the individual process rates work in concert to reproduce the behavior of McSnow in an ODE sense. This is by no means trivial as shown for example by SR20 for warm-rain microphysics. Second, the ML-based P3-like microphysics provides reasonable results for a 3d idealized squall line simulation with ICON. It can in some aspects improve over the SB two-moment scheme in that it produces a more realistic-looking extended stratiform region with a secondary maximum in radar reflectivity. The ML-based scheme achieves this by predicting physically plausible rime mass and rime density of snow and graupel and corresponding sedimentation velocities. The ML-based P3-like scheme is even able to predict the wet growth regime of graupel within the convective core. Third, the ML-based scheme has been applied in a realistic forecast scenario with ICON on a 2 km grid to predict the evolution of a mesoscale convective system. In this case, the ML-based scheme runs stably over a large spatial domain and provides a reasonable representation of the cloud microphysics. Unfortunately, it is not able to improve over the SB two-moment scheme in the chosen case. Most likely, the microphysics is simply not the limiting factor for the forecast quality of this mesoscale convective system, but instead, other model components are relevant as well and would have to be improved.

In contrast to classic bulk microphysics parameterizations, the ML-based P3-like scheme does not explicitly make assumptions regarding particle geometries or particle size distributions. All this is learned from the McSnow simulations in a parameter-free way. This makes the ML approach flexible, but it requires that additional neural nets are trained for diagnostics like radar reflectivity or effective radius. For more complex diagnostics like polarimetric radar variables, which are very challenging for conventional bulk microphysics schemes, the ML approach could be promising, though.

The ML approach chosen here has some disadvantages. First, the ML-based scheme ended up being computationally expensive. On one hand, simply because we decided to build a very complicated scheme with 23 prognostic variables. On the other hand, the implementation with individual neural nets for each physical process that have to be evaluated at each grid point and each time step is rather inefficient. It should be possible to subsequently build an emulator of the ML-based scheme that would overcome these deficiencies. For example, by having fewer neural networks and taking model columns as input instead of individual grid points. Even the calculation of process tendencies can be questioned and instead, a direct mapping of state variables from one time step to the next could be implemented.

Second, the training data is based on rather simplistic box model simulations, which does not make full use of the super-particle model McSnow. McSnow can in principle be applied in 2d or even 3d ICON simulations. Hence, the current ML-based scheme could be used as a baseline model and more training data from 2d and 3d McSnow simulations could further improve the realism of the microphysical processes and their interaction.

A third and more general issue of such ML-based schemes is that they do not allow much a posteriori tuning of the model. Any change in the microphysical assumptions like basic particle geometry or sticking or collision efficiencies, for example, would have to be done in McSnow. Then the full ML workflow has to be repeated including the production of the training data. This makes sensitivity studies to explore model uncertainties very time-consuming. In practice, NWP or climate models do require some a posteriori tuning to balance different physical processes and their biases. Common tuning parameters like intercept parameters of the particle size distribution, terminal fall velocity, or particle geometries cannot easily be modified in ML-based schemes. Changing the bulk sedimentation velocity by a constant factor is possible, but would make the scheme inconsistent. This issue could only be overcome if the ML-based model could be further trained and improved within the atmospheric model itself. Preferably such an online training would be done with actual observations as part of a data assimilation system. First steps toward such an online training capability for ICON are currently being implemented at DWD.

Figure 1. Vertical profiles of mass densities of the various particle classes of the ML-based P3-like scheme. Shown is McSnow output (solid), the ODE solution using training data (dotted) and the ODE solution of the ML-based scheme (dashed). Shown is a simulation with $h_1 = 1500$ m, $h_2 = 3000$ m, $r_c = 15$ μm, $N_i = 20$ dm^{-3}, $Q_i = 0.2$ g cm^{-3}, $Q_c = 0.1$ g cm^{-3} and $S_i = 0$.

Figure 2. As Figure 1, but for number densities.

Figure 3. Vertical profiles of particle properties of rimed snow. Shown is McSnow output (solid), the ODE solution using training data (dotted), and the ODE solution of the ML-based scheme (dashed).

Figure 4. Vertical cross section of radar reflectivity dBZ for the SB two-moment scheme (left) and the ML-based P3-like scheme (right). Shown are averages along the y-direction after 300 min simulation time.

Figure 5. Vertical cross sections of rime fraction (left) and rime density (right), of all hydrometeors (top), rimed snow (center) and graupel (bottom) as predicted by the ML-based P3-like scheme. Shown are averages along the y-direction after 300 min simulation time.

Figure 6. Vertical cross section of the liquid water fraction of rimed snow (left) and graupel (right) of the ML-based P3-like scheme. Shown are averages along the y-direction after 300 min simulation time.

Figure 7. Column maximum radar reflectivity for 19 May 2022, 13:30 UTC for a central European region. Shown is the Opera composite (left), the ICON-D2 simulation using the SB two-moment microphysics (center) and ICON using the ML-based P3-like two-moment scheme (right).

Table 4. Overview of ML models of ice microphysical processes. Mean absolute error (MAE) and mean squared error (MSE) are shown for the normalized testing data and are dimensionless. All ML models use ReLU activation and apply the Adam optimizer.

No.	physical process	predictors (features)	output (labels)	NN size	training samples	testing samples	MAE	MSE
1	deposition/sublimation of unrimed ice	q_i, n_i, T, ρ	$\partial_t q_i/S_i$	8x1	1,234,870	263,676	0.0130	0.0036
2	deposition/sublimation of unrimed snow	q_s, n_s, T, ρ	$\partial_t q_s/S_i$	6x1	1,279,759	273,979	0.0307	0.0059
3	deposition/sublimation of graupel	q_g, n_g, ψ_g, ϕ_g, T, ρ	$\partial_t q_g/S_i$	6x1	464,361	99,118	0.0491	0.0253
4	deposition/sublimation of rimed ice	q_{ri}, n_{ri}, ψ_{ri}, ϕ_{ri}, T, ρ	$\partial_t q_{\mathrm{ri}}/S_i$	6x1	66,490	14,288	0.139	0.0987
5	deposition/sublimation of rimed snow	q_{rs}, n_{rs}, ψ_{rs}, ϕ_{rs}, T, ρ	$\partial_t q_{\mathrm{rs}}/S_i$	6x1	263,859	57,038	0.136	0.130
6	selfcollection of unrimed ice	q_i, n_i, T, ρ	$\partial_t q_i$, $\partial_t n_i$	16x2	1,212,134	259,225	0.256	0.124
7	selfcollection of unrimed snow	q_s, n_s, T, ρ	$\partial_t n_i$	32x2	1,135,918	243,714	0.104	0.023
8	aggregation of unrimed ice and unrimed snow	q_i, n_i, q_s, n_s, T, ρ	$\partial_t n_i$, $\partial_t q_i$	16x2	1,497,110	320,229	0.073	0.0178
9	selfcollection of rimed ice	q_{ri}, n_{ri}, ψ_{ri}, ϕ_{ri}, T, ρ	$\partial_t n_{\mathrm{ri}}$, $\partial_t q_{\mathrm{ri}}$, $\partial_t \psi_{\mathrm{ri}}$, $\partial_t \phi_{\mathrm{ri}}$	16x2	13,759	2,971	0.307	0.173
10	selfcollection of rimed snow	q_{rs}, n_{rs}, ψ_{rs}, ϕ_{rs}, $\hat{\ell}_{\mathrm{rs}}$	$\partial_t n_{\mathrm{rs}}$	16x2	1,425,976	305,885	0.0829	0.0177
11	aggregation of unrimed ice and rimed snow	q_i, n_i, q_{rs}, n_{rs}, ψ_{rs}, ϕ_{rs}, T, ρ	$\partial_t n_i$, $\partial_t q_i$	16x2	337,354	72,545	0.0688	0.0170
12	aggregation of unrimed snow and rimed snow	q_s, n_s, q_{rs}, n_{rs}, ψ_{rs}, ϕ_{rs}, T, ρ	$\partial_t n_s$, $\partial_t q_s$	16x2	238,671	51,669	0.0823	0.0192
13	aggregation of rimed ice and rimed snow	q_{ri}, n_{ri}, ψ_{ri}, ϕ_{ri}, q_{rs}, n_{rs}, ψ_{rs}, ϕ_{rs}, T, ρ	$\partial_t n_{\mathrm{ri}}$, $\partial_t q_{\mathrm{ri}}$, $\partial_t \psi_{\mathrm{ri}}$, $\partial_t \phi_{\mathrm{ri}}$	16x2	111,188	24,006	0.277	0.180

Table 5. Overview of ML models (continued)

No.	physical process	predictors (features)	output (labels)	NN size	training samples	testing samples	MAE	MSE
14	conversion of ice to rimed ice	q_i, n_i, q_c, r_c, T, ρ	$\partial_t n_{\rm ri}$, $\partial_t q_{\rm ri}$, $\partial_t \psi_{\rm ri}$, $\partial_t \phi_{\rm ri}$	16x2	12,059	2,495	0.236	0.236
15	conversion of snow to rimed snow	q_s, n_s, q_c, r_c, T, ρ	$\partial_t n_{\rm rs}$, $\partial_t q_{\rm rs}$, $\partial_t \psi_{\rm rs}$, $\partial_t \phi_{\rm rs}$	16x2	10,492	2,245	0.206	0.102
16	riming of rimed ice	$q_{\rm ri}$, $n_{\rm ri}$, $\psi_{\rm ri}$, $\phi_{\rm ri}$, q_c, r_c, T, ρ	$\partial_t q_{\rm ri}$, $\partial_t \psi_{\rm ri}$, $\partial_t \phi_{\rm ri}$	16x2	33,505	7,141	0.152	0.0753
17	riming of rimed snow with cloud droplets	$q_{\rm rs}$, $n_{\rm rs}$, $\psi_{\rm rs}$, $\phi_{\rm rs}$, $\hat{\ell}_{\rm rs}$, q_c, r_c, T	$\partial_t q_{\rm rs}$, $\partial_t n_{\rm rs}$, $\partial_t \psi_{\rm rs}$, $\partial_t \phi_{\rm rs}$, $\partial_t \ell_{\rm rs}$	16x2	227,928	49,345	0.114	0.0665
18	riming of graupel with cloud droplets	q_g, n_g, ψ_g, ϕ_g, $\hat{\ell}_g$, q_c, r_c, T, ρ	$\partial_t q_g$, $\partial_t n_g$, $\partial_t \psi_g$, $\partial_t \phi_g$, $\partial_t \ell_g$	16x2	530,385	113,628	0.0966	0.0481
19	internal melting of rimed snow	$q_{\rm rs}$, $n_{\rm rs}$, $\psi_{\rm rs}$, $\phi_{\rm rs}$, $\hat{\ell}_{\rm rs}$	$\partial_t \ell_{\rm rs}/(T - T_3)$	16x2	810,109	173,193	0.0227	0.0027
20	melting of rimed snow to rain	$q_{\rm rs}$, $n_{\rm rs}$, $\psi_{\rm rs}$, $\phi_{\rm rs}$, $\hat{\ell}_{\rm rs}$, T	$\partial_t \ell_{\rm rs}$, $\partial_t n_{\rm rs}$	16x2	418,529	89,765	0.179	0.0901
21	internal melting of graupel	q_g, n_g, ψ_g, ϕ_g, $\hat{\ell}_g$	$\partial_t \ell_g/(T - T_3)$	16x2	558,006	119,939	0.020	0.0020
22	melting of graupel to rain	q_g, n_g, ψ_g, ϕ_g, $\hat{\ell}_g$, T	$\partial_t \ell_g$,$\partial_t n_g$	16x2	311,931	66,764	0.154	0.088
23	internal freezing of graupel	q_g, n_g, ψ_g, ϕ_g, $\hat{\ell}_g$, q_c, r_c, q_r, n_r	$\partial_t \ell_g/(T_3 - T)$	16x2	420,439	90,226	0.151	0.088
24	collection of rain by graupel	q_g, n_g, ψ_g, ϕ_g, $\hat{\ell}_g$, q_r, n_r, T, ρ	$\partial_t \ell_g$	16x2	671,660	144,110	0.124	0.0624
25	ice multiplication for graupel	q_g, n_g, ψ_g, ϕ_g, $\hat{\ell}_g$, T, ρ	$\partial_t n_i$	16x2	118,906	25,574	0.0894	0.0272
26	ice multiplication for rimed snow	$q_{\rm rs}$, $n_{\rm rs}$, $\psi_{\rm rs}$, $\phi_{\rm rs}$, $\hat{\ell}_{\rm rs}$, T, ρ	$\partial_t n_i$	16x2	68,062	14,673	0.120	0.0467
27	ice multiplication for rimed ice	$q_{\rm ri}$, $n_{\rm ri}$, $\psi_{\rm ri}$, $\phi_{\rm ri}$, T, ρ	$\partial_t n_i$	16x2	9,819	2,068	0.109	0.0330
28	sedimentation of unrimed ice	q_i, n_i, T, ρ	v_q, v_n	16x2	979,504	209,552	0.078	0.0163

Table 6. Overview of ML models (continued)

No.	physical process	predictors (features)	output (labels)	NN size	training samples	testing samples	MAE	MSE
29	sedimentation of unrimed snow	q_s, n_s, T, ρ	v_q, v_n	16x1	1,707,255	365,700	0.076	0.020
30	sedimentation of rimed ice	q_{ri}, n_{ri}, ψ_{ri}, ϕ_{ri}, T, ρ	v_q, v_n, v_ψ	32x2	152,090	32,373	0.0863	0.0177
31	sedimentation of rimed snow	q_{rs}, n_{rs}, ψ_{rs}, ϕ_{rs}, $\hat{\ell}_{\mathrm{rs}}$, T, ρ	v_q, v_n, v_ψ, v_ℓ	32x2	1,500,619	321,431	0.0732	0.0194
32	sedimentation of graupel	q_g, n_g, ψ_g, ϕ_g, $\hat{\ell}_g$, ρ	v_q, v_n, v_ψ, v_ℓ	32x2	1,859,964	398,612	0.0562	0.0113
33	radar reflectivity of graupel	q_g, n_g, ψ_g, ϕ_g	z_g	16x2	416,676	89,544	0.0713	0.0118
34	radar reflectivity of rimed snow	q_{rs}, n_{rs}, ψ_{rs}, ϕ_{rs}	z_{rs}	16x2	1,076,404	230,621	0.084	0.017
35	radar reflectivity of rimed ice	q_{ri}, n_{ri}, ψ_{ri}, ϕ_{ri}	z_{ri}	16x2	161,476	34,489	0.051	0.0051
36	radar reflectivity of unrimed ice	q_i, n_i, T, ρ	z_i	16x2	1,122,035	240,143	0.045	0.0065
37	radar reflectivity of unrimed snow	q_s, n_s, T, ρ	z_s	16x2	1,158,769	248,745	0.049	0.0075
38	effective radius of unrimed ice	q_i, n_i, T	$r_{eff,i}$	16x2	1,666,618	356,002	0.0460	0.0084
39	effective radius of unrimed snow	q_s, n_s, T	$r_{eff,s}$	16x2	1,120,739	240,259	0.0468	0.0084
40	effective radius of rimed ice	q_{ri}, n_{ri}, ψ_{ri}, ϕ_{ri}	$r_{eff,\mathrm{ri}}$	16x2	182,967	39,466	0.1409	0.0950
41	effective radius of rimed snow	q_{rs}, n_{rs}, ψ_{rs}, ϕ_{rs}, $\hat{\ell}_{\mathrm{rs}}$	$r_{eff,\mathrm{rs}}$	16x2	1,497,472	320,385	0.0519	0.0323
42	effective radius of graupel	q_g, n_g, ψ_g, ϕ_g, $\hat{\ell}_g$	$r_{eff,g}$	16x2	1,861,272	398,483	0.0381	0.0063
	classifier network						accuracy	
43	selfcollection of unrimed snow	q_s, n_s, n_i, T, ρ	P_{self}	16x2	1,719,206	368,237	0.938	

Appendix A Sticking efficiency in McSnow

There are some observations and laboratory measurements of the sticking efficiency of ice crystals as a function of temperature (Hosler & Hallgren, 1960; Mitchell, 1988; Kajikawa & Heymsfield, 1989; Connolly et al., 2012) but for graupel-graupel collisions or partially rimed snowflakes, the authors are not aware of any measurements. Phillips et al. (2015) discuss the dependency of the sticking efficiency on the collision kinetic energy and, hence, provide a theoretical framework to explain the decrease of the sticking efficiency with increasing degree of riming. Due to the lack of data, a consistent physically-based parameterization is beyond the scope of this study. Nevertheless, a reasonable and continuous description is required to generate meaningful training data for the P3 approach. In the current study, we use the degree of riming defined as

$$\xi = \frac{m_r + m_f}{m_r + m_f + m_i + m_\ell} = \frac{m_r + m_f}{m_{\text{tot}}} \tag{A1}$$

with the rime mass m_r, the frozen mass m_f, the ice (crystal) mass m_i, the liquid mass m_ℓ and the total particle mass m_{tot}.

The following ad-hoc parameterization for the sticking efficiency E_s of two particles a and b has been used in the McSnow simulations:

$$E_s = \begin{cases} E_i, & \text{for } \xi_a + \xi_b < \xi_1 \\ E_g, & \text{for } \xi_a + \xi_b > \xi_2 \\ E_i \frac{\xi_a + \xi_b - \xi_2}{\xi_1 - \xi_2} + E_g \frac{\xi_a + \xi_b - \xi_1}{\xi_2 - \xi_1} & \end{cases} \tag{A2}$$

with $\xi_1 = 0.01$ and $\xi_2 = 0.9$. Here E_i is the temperature-dependent piecewise linear sticking efficiency of unrimed crystals

$$E_i = \begin{cases} 0.07, & \text{for } T_c \geq 0\,^\circ\text{C} \\ -0.005\,(T_c + 10) + 0.12, & \text{for } 0\,^\circ\text{C} > T_c \geq -10\,^\circ\text{C} \\ -0.040\,(T_c + 15) + 0.32, & \text{for } -10\,^\circ\text{C} > T_c \geq -15\,^\circ\text{C} \\ 0.050\,(T_c + 20) + 0.14, & \text{for } -15\,^\circ\text{C} > T_c \geq -20\,^\circ\text{C} \\ 0.0025\,(T_c + 40) + 0.04, & \text{for } -20\,^\circ\text{C} > T_c \geq -40\,^\circ\text{C} \\ 0.02, & \text{for } -40\,^\circ\text{C} > T_c \end{cases} \tag{A3}$$

where T_c is the temperature in degrees Celsius. This formula is largely based on data of Connolly et al. (2012) as shown by their Figure 14, but we intentionally decided on the lower range of those measurements. This choice leads to a more pronounced trailing stratiform region of the idealized squall line.

$$E_g = 0.01 \tag{A4}$$

is the sticking efficiency of graupel-graupel collisions. This sticking efficiency is only applied to ice particles that have no liquid water at the particle surface. For melting particles and in the wet growth regime the sticking efficiency is set to one in McSnow. Both, E_i and E_g, are often used as tuning parameters in cloud simulations. See, for example, the discussion in Karrer et al. (2021) for the sticking efficiency of aggregates.

Appendix B Some implementation details in ICON

To be able to run the ML-based P3-like scheme stably in ICON, a few constraints are necessary. The values of cloud liquid water content and mean cloud particle radius are limited to a range not far beyond the training data. For cloud liquid water this is only an upper bound of $20{\times}10^{-3}$ kg m^{-3}, which should rarely be reached. The cloud droplet radius is forced to be within 5-30 μm when passed to the neural networks. For the sedimentation velocities, upper and lower limits are imposed as always in the SB two-moment scheme. In addition, it is enforced that the sedimentation velocity of mass is larger than that for number. This is done by simply using the larger of the two velocities provided by the neural net for mass, whereas the smaller one is used for number. For the sink term of cloud droplet or raindrop number by riming the mean mass is assumed to be constant as the NNs do not yet provide the information about the size of the collected liquid drops. All microphysical processes except diffusional growth (deposition/sublimation) are only calculated if the mass content of the hydrometeor class exceeds 1×10^{-9} kg m^{-3}. For self-collection of rimed snow lower limits of 1×10^{-5} kg m^{-3} and 10 m^{-3} have to be exceeded for mass and number density, respectively. Melting is only calculated a long as the liquid fraction is below 0.99, then the remaining mass is instantly converted to rain. Conversion of rimed snow to rain by melting only happens if the bulk liquid fraction exceeds 0.3. All processes that are only relevant in the mixed-phase regime, are only calculated in the temperature range 236 K to 273 K. Ice multiplication is restricted to 265.9 K and 270.1 K and limited to 100 m^{-3} s^{-1}. The ML-based P3-like scheme is implemented in ICON as an extension of the SB two-moment scheme. The SB two-moment microphysics additionally enforces the particle sizes of each hydrometeor class to be within a physically meaningful range.

References

Aligo, E. A., Ferrier, B., & Carley, J. R. (2018). Modified NAM microphysics for forecasts of deep convective storms. *Mon. Wea. Rev.*, *146*(12), 4115–4153.

Arnold, C., Sharma, S., Weigel, T., & Greenberg, D. (2023). Efficient and Stable Coupling of the SuperdropNet Deep Learning-based Cloud Microphysics (v0.1.0) to the ICON Climate and Weather Model (v2.6.5). *EGUsphere*, *2023*, 1–17.

Barklie, R., & Gokhale, N. (1959). *The freezing of supercooled water drops.* (MW-30)

Bauer, P., Dueben, P. D., Hoefler, T., Quintino, T., Schulthess, T. C., & Wedi, N. P. (2021). The digital revolution of Earth-system science. *Nature Computational Science*, *1*(2), 104–113.

Bauer, P., Stevens, B., & Hazeleger, W. (2021). A digital twin of Earth for the green transition. *Nature Climate Change*, *11*(2), 80–83.

Böhm, J. P. (1992a). A general hydrodynamic theory for mixed-phase microphysics. Part I: drag and fall speed of hydrometeors. *Atmos. Res.*, *27*(4), 253-274.

Böhm, J. P. (1992b). A general hydrodynamic theory for mixed-phase microphysics. Part II: Collision kernels for coalescence. *Atmos. Res.*, *27*(4), 275–290.

Böhm, J. P. (1992c). A general hydrodynamic theory for mixed-phase microphysics. Part III: Riming and aggregation. *Atmos. Res.*, *28*(2), 103–123.

Böhm, J. P. (1994). Theoretical collision efficiencies for riming and aerosol impaction. *Atmos. Res.*, *32*(1-4), 171–187.

Böhm, J. P. (1999). Revision and clarification of "A general hydrodynamic theory for mixed-phase microphysics". *Atmos. Res.*, *52*(3), 167–176.

Böhm, J. P. (2004). Reply to Comment on "Revision and clarification of 'A general hydrodynamic theory for mixed-phase microphysics'". *Atmos. Res.*, *69*(3-4), 289–293.

Brdar, S., & Seifert, A. (2018). McSnow: A Monte-Carlo particle model for riming and aggregation of ice particles in a multidimensional microphysical phase space. *J. Adv. in Modeling Earth Systems*, *10*(1), 187-206. doi: https://doi.org/10.1002/2017MS001167

Bringi, V., Seifert, A., Wu, W., Thurai, M., Huang, G.-J., & Siewert, C. (2020). Hurricane dorian outer rain band observations and 1d particle model simulations: A case study. *Atmosphere*, *11*(8), 879. doi: 10.3390/atmos11080879

Chantry, M., Hatfield, S., Dueben, P., Polichtchouk, I., & Palmer, T. (2021). Machine learning emulation of gravity wave drag in numerical weather forecasting. *J. Adv. in Modeling Earth Systems*, *13*(7), e2021MS002477. doi: https://doi.org/10.1029/2021MS002477

Cholette, M., Milbrandt, J. A., Morrison, H., Paquin-Ricard, D., & Jacques, D. (2023). Combining triple-moment ice with prognostic liquid fraction in the P3 microphysics scheme: Impacts on a simulated squall line. *J. Adv. in Modeling Earth Systems*, *15*(4), e2022MS003328. (e2022MS003328 2022MS003328)

Cholette, M., Morrison, H., Milbrandt, J. A., & Thériault, J. M. (2019). Parameterization of the bulk liquid fraction on mixed-phase particles in the predicted particle properties (P3) scheme: Description and idealized simulations. *J. Atmos. Sci.*, *76*(2), 561 - 582. doi: 10.1175/JAS-D-18-0278.1

Cholette, M., Thériault, J. M., Milbrandt, J. A., & Morrison, H. (2020). Impacts of predicting the liquid fraction of mixed-phase particles on the simulation of an extreme freezing rain event: The 1998 north american ice storm. *Mon. Wea. Rev.*, *148*(9), 3799 - 3823. doi: 10.1175/MWR-D-20-0026.1

Cober, S. G., & List, R. (1993). Measurements of the heat and mass transfer parameters characterizing conical graupel growth. *J. Atmos. Sci.*, *50*(11), 1591–1609.

Connolly, P., Emersic, C., & Field, P. (2012). A laboratory investigation into the aggregation efficiency of small ice crystals. *Atmos. Chem. Phys.*, *12*(4), 2055–2076.

Feingold, G., Walko, R., Stevens, B., & Cotton, W. (1998). Simulations of marine stratocumulus using a new microphysical parameterization scheme. *Atmos. Res.*, *47*, 505–528.

Field, P. R., Lawson, R. P., Brown, P. R., Lloyd, G., Westbrook, C., Moisseev, D., ... others (2017). Secondary ice production: Current state of the science and recommendations for the future. *Meteorological Monographs*, *58*, 7–1.

Gettelman, A., Gagne, D. J., Chen, C.-C., Christensen, M., Lebo, Z., Morrison, H., & Gantos, G. (2021). Machine learning the warm rain process. *J. Adv. in Modeling Earth Systems*, *13*(2), e2020MS002268.

Hallett, J., & Mossop, S. (1974). Production of secondary ice particles during the riming process. *Nature*, *249*(5452), 26–28.

Hosler, C., & Hallgren, R. (1960). The aggregation of small ice crystals. *Discussions of the Faraday Society*, *30*, 200–207.

Jakob, C. (2010). Accelerating progress in global atmospheric model development through improved parameterizations: Challenges, opportunities, and strategies. *Bull. Am. Met. Soc.*, *91*(7), 869–876.

Kajikawa, M., & Heymsfield, A. J. (1989). Aggregation of ice crystals in cirrus. *J. Atmos. Sci.*, *46*(20), 3108–3121.

Kärcher, B., Hendricks, J., & Lohmann, U. (2006). Physically based parameterization of cirrus cloud formation for use in global atmospheric models. *J. Geophys. Res.*, *111*(D1).

Karrer, M., Seifert, A., Ori, D., & Kneifel, S. (2021). Improving the representation of aggregation in a two-moment microphysical scheme with statistics of multi-frequency doppler radar observations. *Atmos. Chem. Phys.*, *21*(22), 17133–17166.

Khain, A., Beheng, K., Heymsfield, A., Korolev, A., Krichak, S., Levin, Z., … Yano, J.-I. (2015). Representation of microphysical processes in cloud-resolving models: Spectral (bin) microphysics versus bulk parameterization. *Reviews of Geophysics*, *53*(2), 247–322.

Köhler, C. G., & Seifert, A. (2015). Identifying sensitivities for cirrus modelling using a two-moment two-mode bulk microphysics scheme. *Tellus B: Chemical and Physical Meteorology*, *67*(1), 24494.

Lagerquist, R., Turner, D., Ebert-Uphoff, I., Stewart, J., & Hagerty, V. (2021). Using deep learning to emulate and accelerate a radiative transfer model. *J. Atmos. Ocean. Tech.*, *38*(10), 1673–1696.

Meyer, D., Grimmond, S., Dueben, P., Hogan, R., & van Reeuwijk, M. (2022). Machine learning emulation of urban land surface processes. *J. Adv. in Modeling Earth Systems*, *14*(3), e2021MS002744. doi: https://doi.org/10.1029/2021MS002744

Meyer, D., Hogan, R. J., Dueben, P. D., & Mason, S. L. (2022). Machine learning emulation of 3d cloud radiative effects. *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002550. doi: https://doi.org/10.1029/2021MS002550

Milbrandt, J. A., & Morrison, H. (2016). Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part III: Introduction of multiple free categories. *J. Atmos. Sci.*, *73*(3), 975–995.

Milbrandt, J. A., Morrison, H., Dawson II, D. T., & Paukert, M. (2021). A triple-moment representation of ice in the predicted particle properties (P3) microphysics scheme. *J. Atmos. Sci.*, *78*(2), 439–458.

Mitchell, D. L. (1988). Evolution of snow-size spectra in cyclonic storms. Part I: Snow growth by vapor deposition and aggregation. *J. Atmos. Sci.*, *45*(22), 3431–3451.

Morrison, H., & Milbrandt, J. A. (2015). Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part I: Scheme description and idealized tests. *J. Atmos. Sci.*, *72*(1), 287–311.

Morrison, H., Thompson, G., & Tatarskii, V. (2009). Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Wea. Rev.*, *137*(3), 991 - 1007. doi: 10.1175/2008MWR2556.1

Mosimann, L., Weingartner, E., & Waldvogel, A. (1994). An analysis of accreted drop sizes and mass on rimed snow crystals. *J. Atmos. Sci.*, *51*(11), 1548 - 1558. doi: 10.1175/1520-0469(1994)051⟨1548:AAOADS⟩2.0.CO;2

Phillips, V. T., DeMott, P. J., & Andronache, C. (2008). An empirical parameterization of heterogeneous ice nucleation for multiple chemical species of aerosol. *J. Atmos. Sci.*, *65*(9), 2757–2783.

Phillips, V. T., Formenton, M., Bansemer, A., Kudzotsa, I., & Lienert, B. (2015). A parameterization of sticking efficiency for collisions of snow and graupel with ice crystals: Theory and comparison with observations. *J. Atmos. Sci.*, *72*(12), 4885–4902.

Rasmussen, R. M., & Heymsfield, A. J. (1987a). Melting and shedding of graupel and hail. Part II: Sensitivity study. *J. Atmos. Sci.*, *44*(19), 2764–2782.

Rasmussen, R. M., & Heymsfield, A. J. (1987b). Melting and shedding of graupel and hail. Part I: Model physics. *J. Atmos. Sci.*, *44*(19), 2754–2763.

Rasmussen, R. M., Levizzani, V., & Pruppacher, H. (1984a). A wind tunnel and theoretical study of the melting behavior of atmospheric ice particles. II: A theoretical study for frozen drops of radius < 500 μm. *J. Atmos. Sci.*, *41*(3), 374–380.

Rasmussen, R. M., Levizzani, V., & Pruppacher, H. (1984b). A wind tunnel and theoretical study on the melting behavior of atmospheric ice particles: III. Experiment and theory for spherical ice particles of radius $> 500\ \mu$m. *J. Atmos. Sci.*, *41*(3), 381–388.

Reinert, D., Prill, F., Frank, H., Denhard, M., Baldauf, M., Schraff, C., ... Zängl, G. (2023). *DWD database reference for the global and regional ICON and ICON-EPS forecasting system* (Technical Report and Database Description, Version 2.2.2). Deutscher Wetterdienst. Retrieved from https://www.dwd.de/SharedDocs/downloads/DE/modelldokumentationen/nwv/icon/icon_dbbeschr_aktuell.html

Rotunno, R., Klemp, J. B., & Weisman, M. L. (1988). A theory for strong, long-lived squall lines. *J. Atmos. Sci.*, *45*(3), 463–485.

Scheck, L. (2021). A neural network based forward operator for visible satellite images and its adjoint. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *274*, 107841.

Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Periáñez, A., & Potthast, R. (2016). Kilometre-scale ensemble data assimilation for the cosmo model (kenda). *Quart. J. Roy. Met. Soc.*, *142*(696), 1453–1472.

Segal, Y., & Khain, A. (2006). Dependence of droplet concentration on aerosol conditions in different cloud types: Application to droplet concentration parameterization of aerosol conditions. *J. Geophys. Res.*, *111*(D15).

Seifert, A. (2008). On the parameterization of evaporation of raindrops as simulated by a one-dimensional rainshaft model. *J. Atmos. Sci.*, *65*(11), 3608–3619.

Seifert, A., & Beheng, K. D. (2001). A double-moment parameterization for simulating autoconversion, accretion and selfcollection. *Atmos. Res.*, *59-60*, 265–281.

Seifert, A., & Beheng, K. D. (2006). A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 1: Model description. *Meteorol. Atmos. Phys.*, *92*(1-2), 45–66.

Seifert, A., Köhler, C., & Beheng, K. (2012). Aerosol-cloud-precipitation effects over germany as simulated by a convective-scale numerical weather prediction model. *Atmospheric Chemistry and Physics*, *12*(2), 709–725.

Seifert, A., Leinonen, J., Siewert, C., & Kneifel, S. (2019). The geometry of rimed aggregate snowflakes: A modeling study. *J. Adv. in Modeling Earth Systems*, *11* (3), 712-731. doi: https://doi.org/10.1029/2018MS001519

Seifert, A., & Rasp, S. (2020). Potential and limitations of machine learning for modeling warm-rain cloud microphysical processes. *J. Adv. in Modeling Earth Systems*, *12* (12), e2020MS002301.

Shima, S.-i., Kusano, K., Kawano, A., Sugiyama, T., & Kawahara, S. (2009). The super-droplet method for the numerical simulation of clouds and precipitation: A particle-based and probabilistic microphysics model coupled with a non-hydrostatic model. *Quart. J. Roy. Met. Soc.*, *135* (642), 1307–1320.

Siewert, C., & Seifert, A. (2018). McSnow: Particle-based microphysics model in ICON. Poster at ICCARUS (ICON/COSMO/CLM/ART User Seminar), Offenbach, Germany. Retrieved from `https://download.dwd.de/pub/DWD/Forschung_und_Entwicklung/ICCARUS2018_presentations_PDF/Monday/P17M_Siewert.pdf` (last accessed 11/12/2023)

Ukkonen, P. (2022). Exploring pathways to more accurate machine learning emulation of atmospheric radiative transfer. *J. Adv. in Modeling Earth Systems*, *14* (4), e2021MS002875. doi: https://doi.org/10.1029/2021MS002875

Ukkonen, P., Pincus, R., Hogan, R. J., Pagh Nielsen, K., & Kaas, E. (2020). Accelerating radiation computations for dynamical models with targeted machine learning and code optimization. *J. Adv. in Modeling Earth Systems*, *12* (12), e2020MS002226. doi: https://doi.org/10.1029/2020MS002226

Ullrich, R., Hoose, C., Möhler, O., Niemand, M., Wagner, R., Höhler, K., ... Leisner, T. (2017). A new ice nucleation active site parameterization for desert dust and soot. *J. Atmos. Sci.*, *74* (3), 699–717. doi: 10.1175/JAS-D-16-0074.1

Vobig, K., Stephan, K., Blahak, U., Khosravian, K., & Potthast, R. (2021). Targeted covariance inflation for 3d-volume radar reflectivity assimilation with the letkf. *Quart. J. Roy. Met. Soc.*, *147* (740), 3789–3805.

Walko, R. L., Cotton, W. R., Meyers, M., & Harrington, J. (1995). New rams cloud microphysics parameterization part i: the single-moment scheme. *Atmos. Res.*, *38* (1-4), 29–62.

Weisman, M. L., & Klemp, J. B. (1982). The dependence of numerically simulated convective storms on vertical wind shear and buoyancy. *Mon. Wea. Rev.*, *110* (6), 504–520.

Welss, J.-N., Siewert, C., & Seifert, A. (2023). Explicit habit-prediction in the Lagrangian super-particle ice microphysics model McSnow. *Authorea Preprints*. (submitted to J. Adv. in Modeling Earth Systems)

Xue, L., Fan, J., Lebo, Z. J., Wu, W., Morrison, H., Grabowski, W. W., … Rasmussen, R. M. (2017). Idealized simulations of a squall line from the mc3e field campaign applying three bin microphysics schemes: Dynamic and thermodynamic structure. *Mon. Wea. Rev.*, *145*(12), 4789 - 4812. Retrieved from `https://journals.ametsoc.org/view/journals/mwre/145/12/mwr-d-16-0385.1.xml` doi: 10.1175/MWR-D-16-0385.1

Zängl, G., Reinert, D., Rípodas, P., & Baldauf, M. (2015). The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quart. J. Roy. Met. Soc.*, *141*(687), 563–579.

Zeng, Y., Blahak, U., & Jerger, D. (2016). An efficient modular volume-scanning radar forward operator for NWP models: description and coupling to the COSMO model. *Quart. J. Roy. Met. Soc.*, *142*(701), 3234–3256.

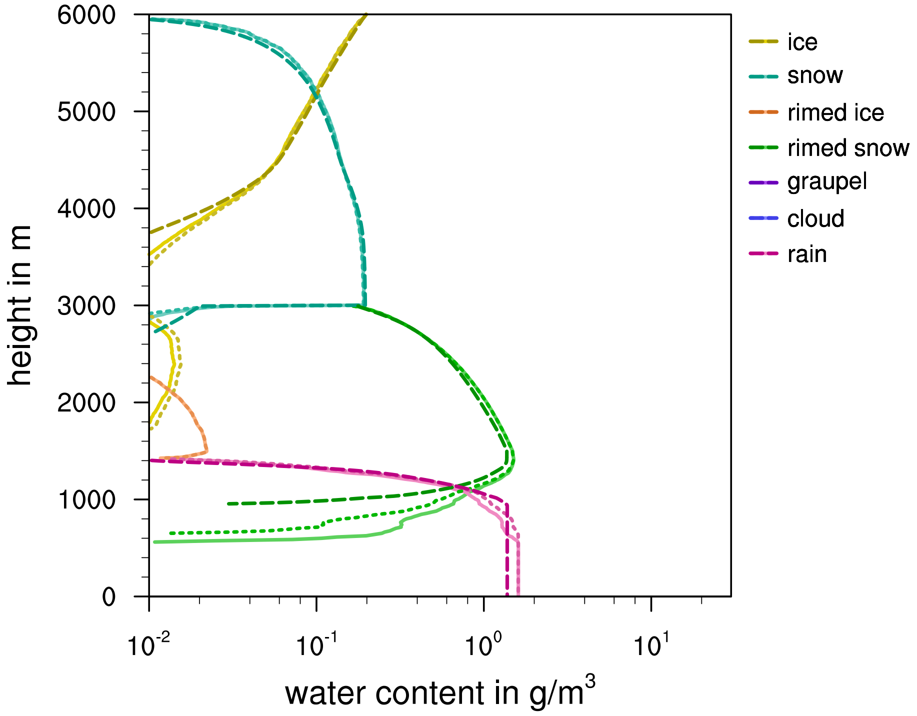
Appendix C Open Research

The training data, the Python notebooks and NCL scripts are provided at Zenodo `doi:10.5281/zenodo.10408950`. The Python notebooks are also publicly accessible at `https://gitlab.com/axelseifert/iceml`. The Zenodo archive does in addition include the Fortran modules containing all the newly developed ICON code for the ML-based P3-like microphysics scheme. The Lagrangian microphysics model McSnow is part of the ICON modeling framework, which is a joint effort of Deutscher Wetterdienst (DWD) and the Max Planck Institute for Meteorology (MPI-M). ICON licenses for scientific use are available at no cost at `https://code.mpimet.mpg.de/projects/iconpublic/`. Subsequently, the access to the McSnow GIT archive can be granted by A.S or C.S.

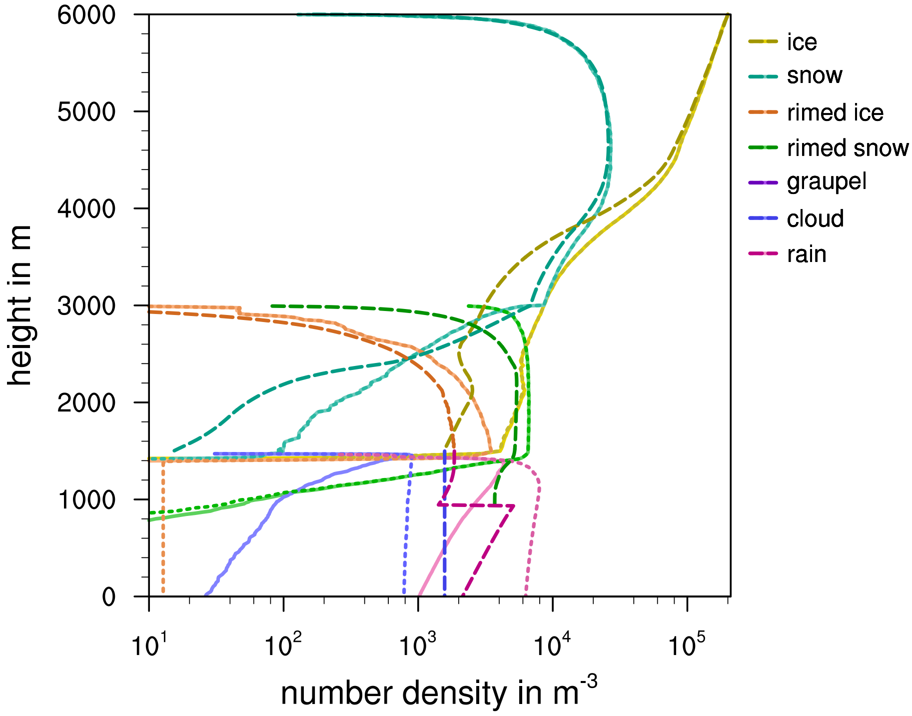
Acknowledgments

The authors thank Jens-Olaf Beismann for help with the implementation and vectorization of the neural nets in ICON and Leonhard Scheck for providing the Fortran code for the multilayer perceptron. The study contributes to the project FRAGILE of the SPP 2115 Fusion of Radar Polarimetry and Numerical Atmospheric Modelling Towards an

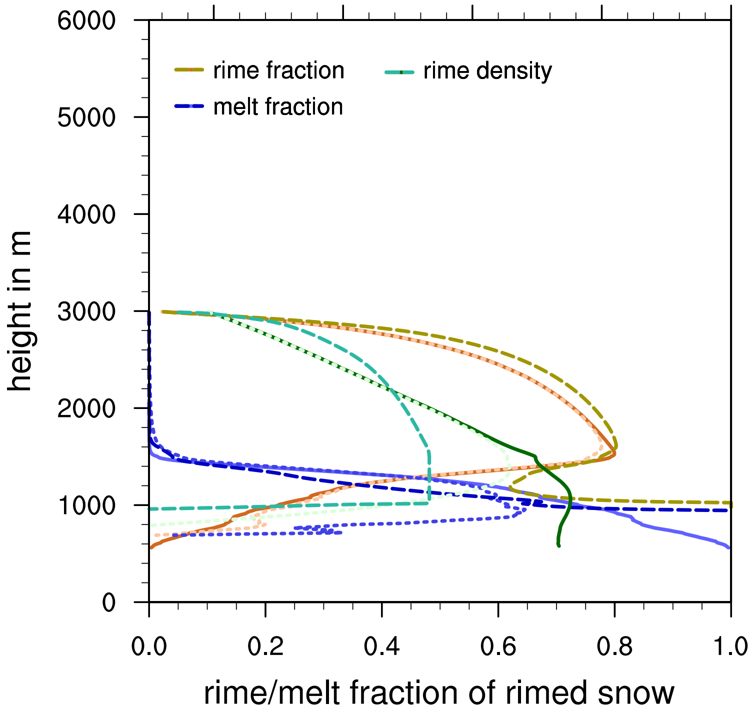
Improved Understanding of Cloud and Precipitation Processes funded by the Deutsche Forschungsgemeinschaft (DFG 492234709).

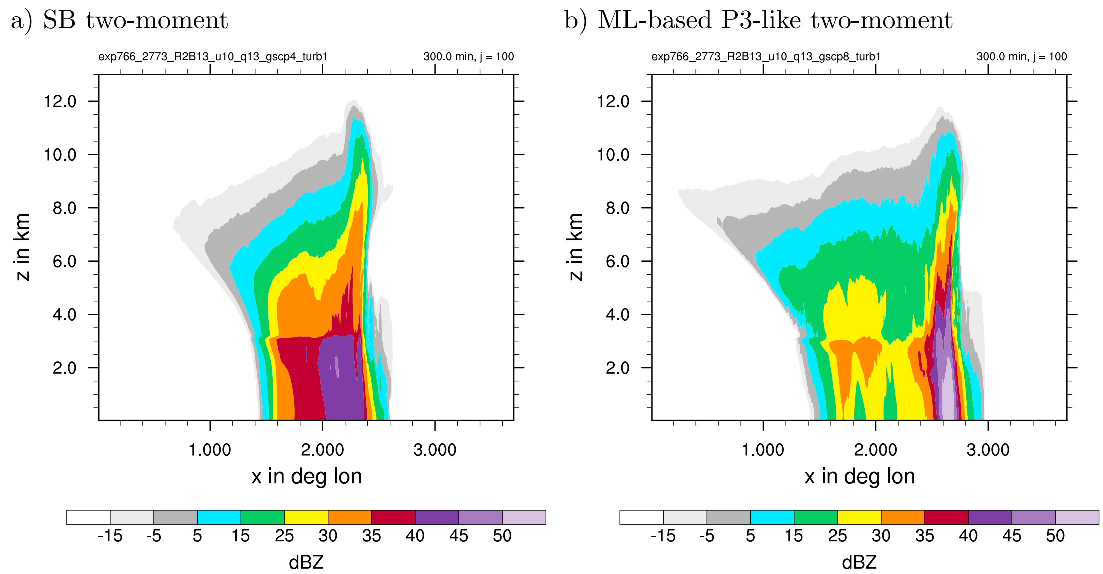


6000
5000
4000
3000
2000
1000
0
height in m
10^{-2}
10^{-1}
10^{0}
10^{1}
water content in g/m^3
ice
snow
rimed ice
rimed snow
graupel
cloud
rain

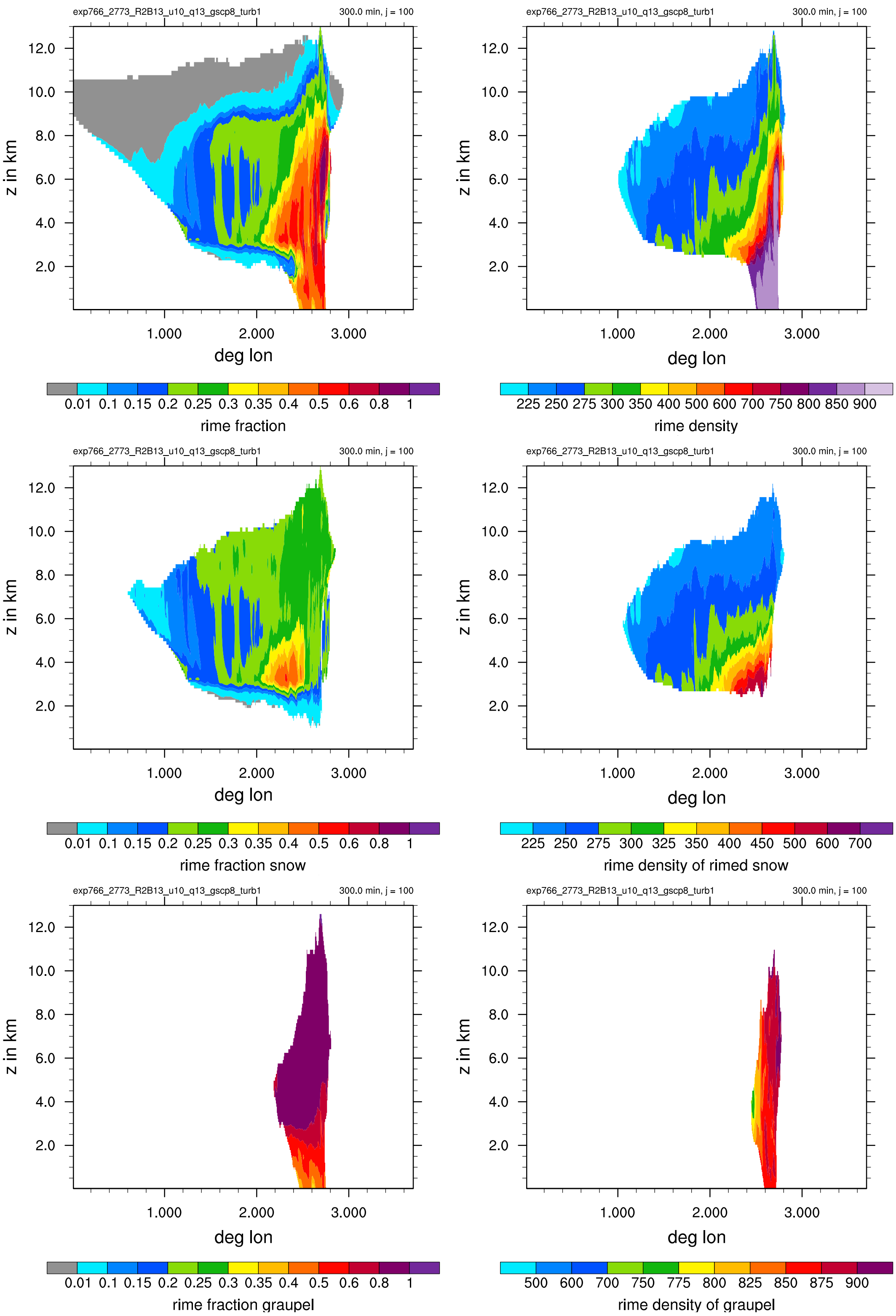


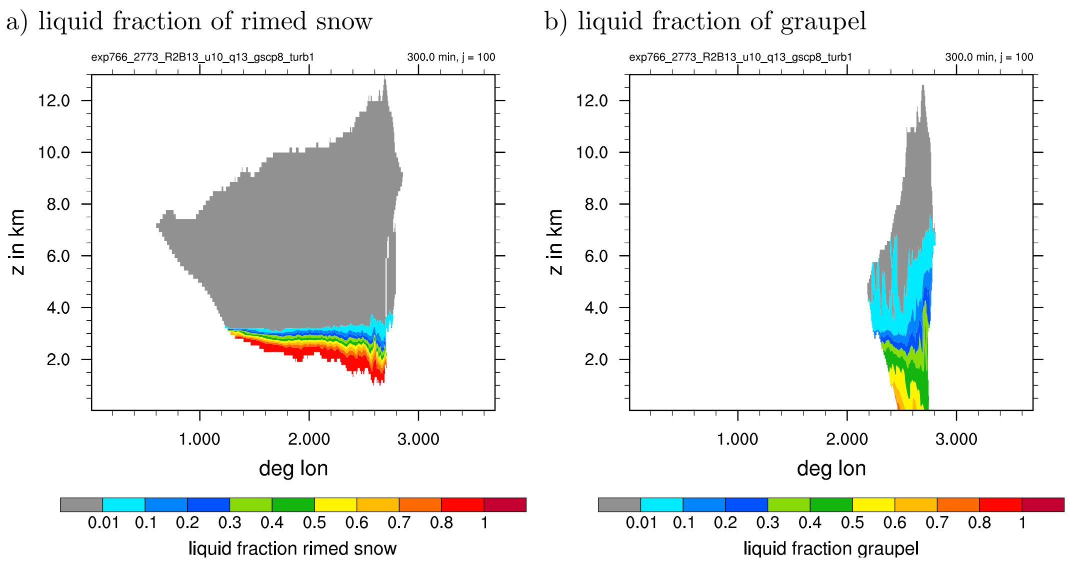
ice
snow
rimed ice
rimed snow
graupel
cloud
rain
height in m
6000
5000
4000
3000
2000
1000
0
10^1
10^2
10^3
10^4
10^5
number density in m^{-3}

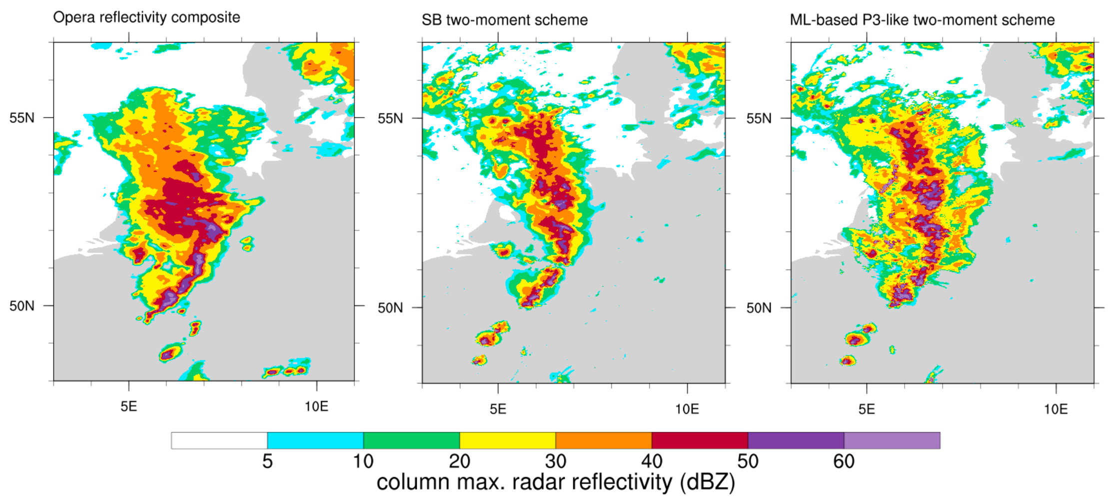

6000
5000
4000
3000
2000
1000
0
height in m
rime fraction
rime density
melt fraction
0.0
0.2
0.4
0.6
0.8
1.0
rime/melt fraction of rimed snow



a) SB two-moment
exp766_2773_R2B13_u10_q13_gscp4_turb1
300.0 min, j = 100
z in km
12.0
10.0
8.0
6.0
4.0
2.0
1.000
2.000
3.000
x in deg lon
-15
-5
5
15
25
30
35
40
45
50
dBZ
b) ML-based P3-like two-moment
exp766_2773_R2B13_u10_q13_gscp8_turb1
300.0 min, j = 100
z in km
12.0
10.0
8.0
6.0
4.0
2.0
1.000
2.000
3.000
x in deg lon
-15
-5
5
15
25
30
35
40
45
50
dBZ


exp766_2773_R2B13_u10_q13_gscp8_turb1
300.0 min, j = 100
z in km
12.0
10.0
8.0
6.0
4.0
2.0
1.000
2.000
3.000
deg lon
0.01 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.5 0.6 0.8 1
rime fraction
225 250 275 300 350 400 500 600 700 750 800 850 900
rime density
rime fraction snow
225 250 275 300 325 350 400 450 500 600 700
rime density of rimed snow
rime fraction graupel
500 600 700 750 775 800 825 850 875 900
rime density of graupel


a) liquid fraction of rimed snow
exp766_2773_R2B13_u10_q13_gscp8_turb1
300.0 min, j = 100
z in km
12.0
10.0
8.0
6.0
4.0
2.0
1.000
2.000
3.000
deg lon
0.01 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 1
liquid fraction rimed snow
b) liquid fraction of graupel
exp766_2773_R2B13_u10_q13_gscp8_turb1
300.0 min, j = 100
z in km
12.0
10.0
8.0
6.0
4.0
2.0
1.000
2.000
3.000
deg lon
0.01 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 1
liquid fraction graupel



Opera reflectivity composite
SB two-moment scheme
ML-based P3-like two-moment scheme
55N
50N
5E
10E
5
10
20
30
40
50
60
column max. radar reflectivity (dBZ)