Using artificial neural networks to predict riming from Doppler cloud radar observations

Teresa Vogl[1], Maximilian Maahn[1], Stefan Kneifel[2], Willi Schimmel[1], Dmitri Moisseev[3,4], and Heike Kalesse-Los[1]

[1]Institute for Meteorology, University of Leipzig, Leipzig, Germany
[2]Institute for Geophysics and Meteorology, University of Cologne, Cologne, Germany
[3]Institute for Atmospheric and Earth System Research/ Physics, Faculty of Science, University of Helsinki, Finland
[4]Finnish Meteorological Institute, Helsinki, Finland

Correspondence: Teresa Vogl (teresa.vogl@uni-leipzig.de)

Abstract. Riming, i.e. the accretion and freezing of ~~SLW~~ supercooled liquid water (SLW) on ice particles in mixed-phase clouds, is an important pathway for precipitation formation. Detecting and quantifying riming using ground-based cloud radar observations is of great interest, however, approaches based on measurements of the mean Doppler velocity (MDV) are unfeasible in convective and orographically influenced cloud systems. Here, we show how artificial neural networks (ANNs) can be used to predict riming using ground-based zenith-pointing cloud radar variables as input features. ANNs are a versatile means to extract relations from labeled data sets, which contain input features along with the expected target values. Training data are extracted from a data set acquired during winter 2014 in Finland, containing both ~~Ka-band~~ Ka- and W-band cloud radar and in-situ observations of snowfall. ANNs are trained separately either on the Ka-band radar or the W-band radar data set. We focus on two configurations of input variables: ANN #1 uses the equivalent radar reflectivity factor (Ze), MDV, the width from left to right edge of the spectrum above the noise floor (spectrum edge width; SEW), and the skewness as input features. ANN #2 only uses Ze, SEW and skewness. The application of these two ANN configurations to case studies from different data sets demonstrates that both are able to predict strong riming (~~riming index = 1~~$FR_{ANN} > 0.7$) and yield low values (~~riming index~~$FR_{ANN} \leq 0.4$) for unrimed snow. In general, the predictions of ANN #1 and ANN #2 are very similar, advocating the capability to predict riming without the use of MDV. ~~It is demonstrated that both ANN setups are able to generalize to W-band radar data.~~ The predictions of both ANNs for a wintertime convective cloud fit coinciding in-situ observations extremely well, suggesting the possibility to predict riming even within convective systems. Application of ANN #2 to an orographic case yields high ~~riming index~~ FR_{ANN} values coinciding with observations of solid graupel particles at the ground.

Copyright statement.

1   Introduction

Mixed-phase clouds are important components of the climate system, because they play a major role both for the radiation budget (Tan et al., 2016) and for the hydrological cycle (Mülmenstädt et al., 2015). In this type of cloud, ice particles and supercooled liquid water (SLW) can coexist in the temperature range between 0 and $-38°$C. The availability of both ice and
SLW allows for riming, i.e. the accretion and freezing of SLW on frozen ice particles. Riming is a key process for ice growth and eventually precipitation formation (Lamb and Verlinde, 2011; Heymsfield et al., 2020). Detecting and quantifying riming using ground-based remote sensing observations is a non-trivial task, but of great interest for several reasons:

Firstly, ice-phase microphysical growth process rates in general are associated with a large uncertainty, posing a major challenge for microphysics schemes in numerical weather forecast and climate models (Morrison et al., 2020). Recently, consider-
able effort has been made to improve the representation of riming in models, ranging from process-oriented approaches (e.g. Seifert et al., 2019) to novel microphysics schemes (e.g. Morrison and Milbrandt, 2015). Deriving the rime mass fraction from remote sensing measurements would enable observation-model comparison studies, which are an important step to evaluate and constrain parameterizations of riming in models. Secondly, those regions where riming occurs in clouds coincide with icing conditions, which pose a hazard to aircraft traffic (Cao et al., 2018). The ability to detect riming conditions e.g. using
ground-based vertically pointing remote sensing instruments would allow for real-time warnings near airports (Serke et al., 2010). Finally, long-term statistics of riming are needed in order to better understand under which conditions riming is taking place in the atmosphere.

In the past, several approaches to retrieve riming from ground-based remote sensing observations have been developed. Riming, at least in its initial stage, increases particle density but not much its size, because the freezing SLW droplets first fill the cav-
ities in the particle structure (Heymsfield, 1982; Seifert et al., 2019). As an effect, its terminal velocity increases (Mosimann et al., 1994; Mosimann, 1995), and also the aspect ratio and backscattering cross-section are changed (Garrett et al., 2015; Leinonen and Szyrmer, 2015). While it is quite challenging to detect riming in polarimetric measurements (Vogel and Fabry, 2018; Li et al., 2018), robust signatures were identified by putting the radar signals of multiple frequencies in relation: Kneifel et al. (2015, 2016) utilized observations of collocated cloud radars with three different wavelengths to pin down fingerprints
of riming and aggregation in the triple-frequency Doppler spectral space. They found that rimed particles can be connected with a combination of low dual-wavelength ratios (DWR) of X and Ka-band radar ($\text{DWR}_{X,Ka} < 3$ dB) and $\text{DWR}_{Ka,W}$ values larger than 3 dB. These signatures are clearer for larger particles. For smaller particles, i.e. small DWR values, this distinction becomes ambiguous. Under these conditions, the fall velocity of the hydrometeors can give an additional constraint. Mason et al. (2018) developed an optimal estimation retrieval to obtain a density factor parameter, which is sensitive to riming, using
observations of mean Doppler velocity (MDV) and $\text{DWR}_{Ka,W}$. Li et al. (2020) developed a snow observation classification with a rimed and an unrimed category, using MDV and $\text{DWR}_{X,Ka}$. Oue et al. (2021) combined MDV and $\text{DWR}_{Ka,W}$ with polarimetric observations and were able to observe even different stages of the riming process. However, only a few sites worldwide are equipped with cloud radars of two or more different frequencies, and correct alignment and volume matching is associated with a certain effort. Another approach which does not rely on observations at multiple wavelengths is based

on MDV only (Mosimann et al., 1994; Mosimann, 1995). Recently, this technique has been further developed by Kneifel and Moisseev (2020), who were able to derive a robust estimate of rime mass fraction using MDV averaged in both spatial and temporal domains over a 100 m / 20 min height-time window. They applied this method to time series of radar measurements at several stations located within the CloudNet (Illingworth et al., 2007) network.

5   While there is an evident correlation between MDV and riming due to the larger density and thus increased fall speed of rimed ice particles, it is not always possible to rely on this relation. The method fails when the assumption that the ~~particle terminal fall velocity equals the measured MDV~~ MDV is equivalent to the particle fall speed in the observation volume does not hold. This is, e.g., the case in convective systems, which can cause persistent up- or downdrafts. Also, in complex terrain, orographically induced waves can shift the observed MDV up or down by several meters per second. These gravity waves can

10   be temporally persistent and consequently not be removed by temporal averaging (Radenz et al., 2021), making MDV-based approaches such as the one by Kneifel and Moisseev (2020) unfeasible. This is unfortunate especially considering the fact that riming plays an important role in the microphysics of convective and orographic systems (Woods et al., 2005; Houze and Medina, 2005). In these cases, other radar variables have to be exploited in search for fingerprints of riming, which allow for a quantitative detection of this process. During riming, multiple hydrometeor populations are present in the same radar observa-

15   tion volume. Due to their different terminal fall velocities, these hydrometeor types result in multiple peaks in the cloud radar Doppler spectra (Kalesse et al., 2016, 2019). A broadening of the spectra is the result, which is e.g. evident in an increase of the spectrum width (the second moment). Another variable which is impacted by riming is the skewness (the fourth moment): Multi-peaked spectra have nonzero aboulte skewness values, zero being the value of a Gaussian distribution. The impact of riming on these other radar variables is, however, not as straightforward as for MDV. For example, riming and aggregation

20   processes may result in similar signatures, because aggregation can also lead to bimodal spectra (e.g. Barrett et al., 2019). Consequently, the often ambiguous information contained in the higher radar moments cannot easily be extracted from the radar variable space using e.g. simple thresholding techniques (Maahn and Löhnert, 2017). More sophisticated methods to derive the relationship between riming and the set of available radar variables are required.

Machine learning (ML) algorithms offer ways to discern relationships from complex data sets. The interest in ML techniques

25   has been increasing rapidly over the past years, and exciting advances have been accomplished in many scientific fields. Also in atmospheric sciences, the use of ML offers a promising path for scientific discoveries (e.g. Seifert and Rasp, 2020). Deep learning is a type of ML, where artificial neural networks (ANNs) are trained to make predictions. Their potential to extract useful relations from ground-based radar observations has been demonstrated in work published e.g. by Luke et al. (2010) and van den Heuvel et al. (2020).

30   In this study, the overall goal is to develop a technique which does not rely on MDV and can consequently be applied to data sets acquired in complex terrain, where orographically induced vertical air motions render MDV unusable. More precisely, our motivation is to derive riming estimates for a data set acquired in Punta Arenas, Chile. This site is strongly influenced by stationary gravity waves, which make the application of MDV-based riming retrievals challenging. At the same time, information about the occurrence of riming over Punta Arenas would be of special interest due to its location in the vicinity of

35   the Southern Ocean and the pristine aerosol conditions encountered there (Foth et al., 2019; Radenz et al., 2021). We assess

in this work how ANNs can be used to predict riming using ground-based zenith-pointing cloud radar measurements as input features. We optimize and train ensembles of ANNs using different combinations of radar variables. These variables include the equivalent radar reflectivity factor (Ze), the MDV, the spectrum width from left to right edge of the spectrum above the noise floor ("spectrum edge width", SEW), and the skewness.

2 Data and methods

2.1 Field experiments

We are using data from the "Biogenic Aerosols – Effects on Clouds and Climate" (BAECC; Petäjä et al., 2016) campaign to train, validate and test ensembles of ANNs. During BAECC, in situ observations of snow were collocated with ~~Ka-band~~ Ka- and W-band radar measurements at a site in Hyytiälä, Finland. The ~~developed models~~ ensembles of trained ANNS are then applied to observations collected during the "TRIple-frequency and Polarimetric radar Experiment for improving process observation of winter precipitation" (TRIPEx-pol; Mróz et al., 2020) in Section 3.2. The motivation to use these data is twofold: Firstly, it allows us to demonstrate that the ANNs are able to generalize to different meteorological conditions and radar settings, and to check whether ~~they can be applied to~~ the predictions are consistent between the Ka- and W-band ~~radar observations as well. Using a different radar frequency could be problematic due to the different sensitivity and scattering properties. A lower or higher radar sensitivity might impact the computed SEW due to its dependence on the noise level, and thus require re-training of the ANNs~~setup. Secondly, the riming retrieval can be additionally compared with expected signatures of riming in triple-frequency radar observations. Another validation study is performed using data obtained at the Leipzig Institute for Meteorology (LIM) on 19 March 2021 in Section 3.3. In this data set, W-band radar data are complemented by ground-based in situ observations of graupel and snowflakes. In Section 3.4, we apply one set of ANNs to a graupel case measured by the same W-band radar during the "Dynamics, Aerosol, Cloud And Precipitation Observations" (DACAPO-PESO) field experiment in Punta Arenas. In this section, each of the four measurement setups is briefly introduced.

2.1.1 The BAECC experiment

The BAECC campaign was a joint effort between the US Department of Energy Atmospheric Radiation Measurement Program (DOE ARM), University of Helsinki, the Finnish Meteorological Institute (FMI), the US National Aeronautics and Space Administration (NASA), and Colorado State University. From February to September 2014, the second ARM mobile facility (AMF2) was deployed at the Station for Measuring Ecosystem-Atmosphere Relations II (SMEAR II) in Hyytiälä, Finland (61° 51′N, 24° 17′E). A detailed description of the setup of the remote sensing and in situ instrumentation can be found in Petäjä et al. (2016). The suite of remote-sensing instruments ~~includes~~ include a Ka-band ARM ~~zenith-pointing~~ Zenith-pointing radar (KAZR), ~~which is a Doppler cloud radar. A ceilometer was located at SMEAR II as well, measuring the cloud base height (CBH)~~and a Marine W-Band ARM Cloud Radar (MWACR), both Doppler cloud radars. Moreover, several ground-based in

situ instruments for measuring microphysical properties of snow were deployed at the site. We use data from the ground-based in situ Precipitation Imaging Package (PIP, Pettersen et al., 2020), a video disdrometer which measures the size and velocity of hydrometeors reaching the surface.

2.1.2   The TRIPEx-pol experiment

The TRIPEx-pol campaign took place from November 2018 until February 2019 at the Jülich Observatory for Cloud Evolution Core Facility, Germany (JOYCE-CF, 50°54′N, 6°25′E, Löhnert et al. (2015)). At that site, vertically pointing pulsed X-, and Ka-band (Görsdorf et al., 2015) systems (manufactured by Metek GmbH) and a Frequency Modulated Continuous Wave ~~vertically-poiting~~ vertically-pointing W-band radar (FMCW, Radiometer Physics GmbH, Küchler et al., 2017) are employed on a roof platform within 10 m distance. Calibration, quality control, matching and resampling procedures described in Dias Neto et al. (2019) were applied to yield high-quality data from all three instruments on the same time-height grid. We are using data from seven days between 24 November 2018 and 10 January 2019 during which thick cloud systems were observed, covering a range of meteorological conditions. The specific days chosen for this analysis are 24 November, 3, 8, 23, and 27 December, and 7 and 10 January, all featuring thick mixed-phase cloud cases, possibly with riming.

2.1.3   LIM roof platform

The University of Leipzig Institute for Meteorology (LIM) remote sensing instrument suite is located on the observatory's roof platform (51°20′N, 12°22′E). This CloudNet site encompasses LIM's 94 GHz Radar (LIMRAD94), and a Video In Situ ~~Snofall~~ Snowfall Sensor (VISSS, Maahn et al., 2021). LIMRAD94 is a W-band FMCW radar very similar to the system which was employed during TRIPEx-pol. The VISSS is an optical observation system for snow particles, measuring 140 frames per second at an optical resolution of 59 μm. While particle size distribution and fall velocity are planned to be avaliable as data products in the future, we here make use of the recorded images of snow particles in a qualitative manner.

2.1.4   The DACAPO-PESO experiment

Since December 2018, the Leipzig Aerosol and Cloud Remote Observations System (LACROS, Bühl et al., 2016) has been employed in Punta Arenas, Chile (53°10′S, 70°56′W). LACROS comprises a suite of active and passive remote sensing instruments, and for the time period from December 2018 to October 2019, LIMRAD94 was employed next to the LACROS shipping container. The measurement site in Punta Arenas is located at the most southerly continental part of South America, where the meteorological conditions are characterized by prevailing strong westerly to northwesterly winds. As the air flow hits the continental mass, it is forced over mountainous orography and at the same time strongly decelerated, resulting in orographic wave motions. From the ground-based radar perspective, persistent up- or downdrafts (~~Lee~~ lee waves), as well as updrafts followed by downdrafts or vice versa can often be observed. The site and the DACAPO-PESO experiment (http://dacapo.tropos.de) have

been described in more detail in Floutsi et al. (2021).

2.1.5 Attenuation corrections

Radar reflectivity of all data used in this study were corrected for attenuation by atmospheric gases, liquid water, melting particles and ice. The Passive and Active Microwave TRAnsfer (PAMTRA) forward operator (Mech et al., 2020) was used in combination with CloudNet products, which are either freely available on the CloudNet data portal (https://cloudnet.fmi.fi/) or were processed locally using the cloudnetpy Python package (Tukiainen et al., 2020): Attenuation due to atmospheric gases, including water vapor, was estimated using PAMTRA and the profiles of temperature and humidity included in the CloudNet model files. For liquid water, we used the measured liquid water path and distributed the mass among the pixels classified as "liquid containing" in the CloudNet classification mask, weighted by the measured reflectivity. We used PAMTRA to obtain the attenuation caused by the mass liquid in the respective height bin, assuming a monodisperse particle size distribution for cloud droplets and an exponential distribution for rain (Mech et al., 2020). If pixels were classified as "melting" in CloudNet, the melting layer attenuation was assumed following Matrosov (2008), who derived relations for Ka- and W-band as a function of rainfall rate. Attenuation due to snow and ice was neglected for Ka-band and estimated according to Protat et al. (2019) for the W-band radars. If the cumulative attenuation correction for a pixel exceeded 10 dBZ, the profile was removed from the analysis.

2.2 Sampling of training data

Masses of individual snow flakes can be retrieved by applying hydrodynamic theory to PIP observations of particle velocity and size (von Lerber et al., 2017). A data set containing observed snow particle number size distributions, along with retrieved particle masses collected between 2014 and 2015 in 5-minute temporal resolution is freely available on GitHub (Moisseev, 2018). Using these retrieved masses m and the maximum dimensions of the observed particles, D_{max}, we first estimated the rime fraction. The mass of unrimed snow, m_{us} is assumed as in Moisseev et al. (2017); Li et al. (2020):

$$m_{us} = \alpha \cdot D_{max}^{\beta} \tag{1}$$

where the values for α and β are $\alpha = 0.0053$ and $\beta = 2.05$ (in cgs units). The rime fraction is then defined as in Kneifel and Moisseev (2020):

$$FR_{PIP} = 1 - \frac{IWC_{us}}{IWC} = 1 - \frac{\int m_{us}(D_{max}) \cdot N(D_{max}) \mathrm{d}D_{max}}{\int m(D_{max}) \cdot N(D_{max}) \mathrm{d}D_{max}} \tag{2}$$

Here, IWC_{us} is the estimated ice water content for unrimed snow having the same $N(D_{max})$ as the observed particles. It is obtained by integrating m_{us} computed from Eq. (1) over the measured particle number size distribution $N(D_{max})$. FR_{PIP} values smaller than zero were removed. In the next step, continuous PIP sampling periods longer than one hour were identified for which KAZR and/ or MWACR observations are available as well. This results in six snow cases between 1 February and 20 March 2014, which are listed in Table 1.

Table 1. Overview of the cases used in the training set, and coincident notable events according to Table 2 in Petäjä et al. (2016)

time period	precipitation type	note
01 February 2014, 01:07 – 01 February 2014, 05:57 UTC	snow	
01 February 2014, 09:07 – 01 February 2014, 15:57 UTC	snow (riming)	
15 February 2014, 20:07 – 16 February 2014, 01:57 UTC	snow (riming)	
21 February 2014, 15:07 – 22 February 2014, 03:27 UTC	snow (riming)/ melting snow	
18 March 2014, 23:07 – 19 March 2014, 19:57 UTC	large aggregates/ riming	no MWACR data
20 March 2014, 15:07 – 20 March 2014, 23:47 UTC	snow/ riming	

During BAECC, a turbulent surface layer was often present, resulting in a broadening of Doppler spectra and impacting their width and other higher-order moments relevant for the training. The following sampling procedure for the training data set was chosen, striving to achieve the best-possible spatio-temporal match between remote sensing and in situ observations, while at the same time avoiding to sample spectra which are strongly impacted by surface induced turbulence: We ~~assume riming to only occur within the cloud and hence FR at CBH is expected to remain constant down to the surface level~~compute the turbulent eddy dissipation rate (EDR) for 5-minute time periods (Maahn et al., 2015) and determine r, the lowest range at which EDR is $< 10^{-3} m^2 s^{-3}$. This threshold was determined empirically in a sensitivity study, in which spectral broadening was simulated using convolution of a spectral broadening term σ_T with measured Doppler spectra. The time ~~x~~ Δt it takes for the particle to travel from ~~the CBH~~ height r to the surface can be estimated using the measured MDV at ~~CBH~~r, making the assumption of temporal and spatial coherence, i.e. that the properties measured at time t ~~at a certain range~~ and r ~~above the radar~~ will be the same as the properties at time ~~$t + x$ at $r + x \cdot \mathrm{MDV}(t,r)$~~$t + \Delta t$ at $r + \Delta t \cdot \mathrm{MDV}(t,r)$. Note that negative MDV values indicate downward motions. A schematic for visualization of the applied sampling technique is shown in Fig. 1. For each PIP measurement, the rime fraction was obtained using Eq. 2. All the KAZR ~~spectra at CBH~~and MWACR spectra at range r, which were matched ~~within~~with a 5-minute ~~window centered at the PIP measurement time stamp~~ PIP measurement were extracted, and Ze, MDV, and skewness were computed. MDV was corrected for air density change as a function of altitude (Vogel and Fabry, 2018). An additional measure of the width of the spectra ~~, which is not weighted by the reflectivity,~~ was computed and stored: The ~~"spectrum edge width" (SEW)~~ SEW is defined as the distance (in $\mathrm{m\,s^{-1}}$) between the left edge and the right edge above the noise threshold. This threshold is the mean noise according to Hildebrand and Sekhon (1974) plus three standard deviations of the noise. SEW offers the advantage that it is, in contrast to the spectrum width, not weighted by the reflectivity, making it more sensitive to broadening of the spectrum due to small peaks, e.g. caused by liquid water. While SEW and skewness are not among the "traditional" radar moments, they are often available; e.g., the ARM MicroARSCL product contains the skewness and edges of the spectra (Kollias et al., 2007; Luke et al., 2008).

The resulting extracted training data set includes ~~59,396~~ 105,115 Ka-band and 209,965 W-band radar observations along with the FR_{PIP} retrieved from the corresponding PIP measurements. It is available on GitHub (https://github.com/ti-vo/BAECC_
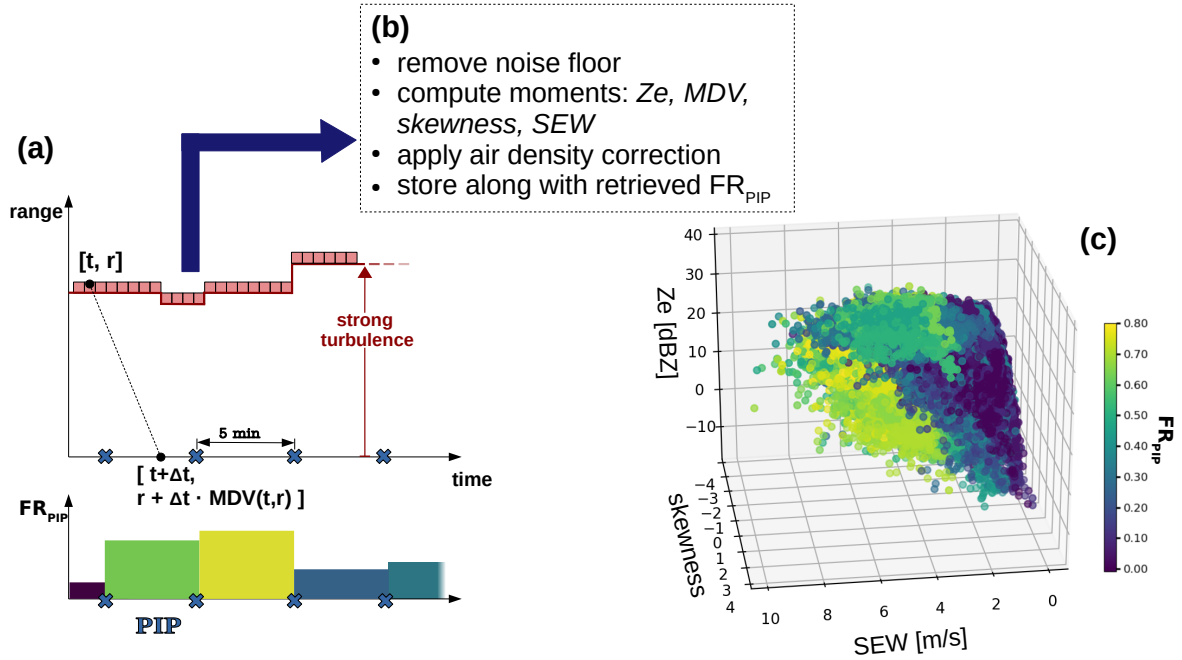
Figure 1. (a) Schematic of the spatio-temporal matching of the in situ (PIP) observations and the cloud radar observations: Radar spectra at ~~cloud base height (measured by a Ceilometer) at~~ time t and range r are assigned to the PIP measurement closest to the time $t + x$. (b) Processing steps applied to each extracted radar Doppler spectrum. (c) 3D plot of three features contained the resulting training data set (equivalent radar reflectivity factor Ze, spectrum edge width SEW, and skewness) colored by the rime mass fraction FR.

features). A 3D plot of three of the contained features (Ze, SEW and skewness) colored by FR_{PIP} is shown in Fig. 1c. The video supplement contains an animated visualization of the same variables. ~~We acknowledge that several sources of uncertainty impact the training data. Assumptions for the density of unrimed snow and the viewing geometry corrected D_{max} are required. Furthermore, errors are introduced by the spatio-temporal matching of radar and PIP observations. Finally, only a limited~~

5 ~~number of cases with high FR, i.e. values > 0.7 were observed during the period when KAZR and PIP were collocated in Hyytiälä. By their nature, high FR cases are rather short-lived, so the duration and therefore length of such observations are rather small. This skewed distribution of target values has impacts on the training of machine learning algorithms and needs to be taken into account carefully.~~

2.3 Machine learning methods

10 The retrieval described in this section reflects a regression problem. Machine learning techniques are used to relate Doppler cloud radar moments and SEW ~~using a fully-connected deep neural networkto predict FR~~to FR_{PIP} using a fully connected neural network. Multiple ANN models are trained to make predictions, given many input (features) and output (target) pairs, with the goal to search for a function that both fits the given data well, and also is able to generalize to new values (Goodfellow

8

et al., 2016). For the steps described in the following, tools provided in the Python library *Sklearn* (Pedregosa et al., 2011) were used.

2.3.1 Data preparation

Radar moments and the SEW extracted from the BAECC data set are used as input features, and ~~PIP-retrieved FR~~FR_{PIP} values are the desired target. One problem emerges because only a limited number of cases with high FR, i.e. values > 0.7 were observed during the period when KAZR, MWACR and PIP were collocated in Hyytiälä. By their nature, high FR cases are rather short-lived, so the duration and therefore length of such observations are rather small. This skewed distribution of target values has impacts on the training of machine learning algorithms and needs to be taken into account carefully. ANN models will primarily be trained to predict values in the range where most of the observations lie. In order to ensure that trained models also succeed to predict high values of FR, which are of special interest, the training data set is resampled using the synthetic minority oversampling technique for regression problems (SMOTER, Torgo et al., 2013). SMOTER is an algorithm which oversamples rare cases and undersamples frequent cases in the training data set, leading to a more balanced distribution. For many ML applications, standardization of the input data set is a common requirement. In order to avoid effects caused by outliers, it is advantageous to use a robust scaler, which scales each of the input features according to the inter-quartile range (IQR) and removes the median. The same robust scaler which is defined using the training data set will also be applied later on to new data being sent through the model.

After scaling, the labeled data are split into two parts, one training and validation set (90%), and retaining 10% of the data for the testing phase. The choice of these ratios is subjective and associated with a tradeoff between learning and the assessment of the generalization ability.

2.3.2 ML model specifications

In our study, we are using a multilayer perceptron (MLP), which is composed of one input layer, at least one hidden layer, and one output layer. In this section, the hyperparameters, which determine the network's architecture and the training process, are discussed.

The input layer consists of one neuron for each input feature, and the output layer yields the output value(s). In between the input and output layers, one or more hidden layers are defined. Each hidden layer contains a certain number of neurons, which transform the values from the previous layer and then apply a non-linear (rectified linear unit; $f(x) = max(0, x)$) activation function. In a fully connected model, a given neuron is connected to every neuron in the previous layer. During training, an optimizer iteratively adjusts the weights of the model, until a minimum in error ("loss") is reached. This error is defined by the loss function, in our case the squared error. We are using the "Adaptive Moment Estimation" (Adam) optimizer, a stochastic gradient-based method (Kingma and Ba, 2017), which requires computing the gradient of the loss function with respect to the model parameters by a back-propagation algorithm. The learning rate controls the step-size for updating the weights, and is in our case set to a constant value of 0.001.

The number of hidden layers (which is also referred to as the layer depth) and respective number of neurons are hyperparameters, which require tuning in the validation phase. In Goodfellow et al. (2014), plots of accuracy vs.layer depth are used to determine the optimal ANN architecture. Following this concept, we plotted the root mean squared error (RMSE) of the validation data set predictions and targets vs. the number of neurons/ hidden layers (not shown) to find an architecture which

5   is simple yet accurate. For the limited number of input features used in our approach, a relatively simple setup with one hidden layer and two neurons fulfills these requirements.

Here, we use a k-fold cross validation (k-fold cv) approach with ~~five~~ three folds. This means that the training and validation set is split into ~~five~~ three equally-sized chunks (folds). Each of the folds is once used to test the model while the remaining four folds are used to train the model. This method enables us to compute the mean error of the ~~five~~ three folds, which is a

10   more reliable measure of the ANN performance than the error obtained for one split between training/ validation set only. In addition, this approach results in ~~five~~ three trained models, which can be applied as an ensemble to yield an average prediction and a ~~variance~~spread, which can (at least to some degree) be regarded as an uncertainty estimate of the prediction. In this study, we are using three different parameter combinations as input features to the ANNs:

1. ANN #0: Ze, MDV

15   2. ANN #1: Ze, MDV, SEW and skewness

3. ANN #2: Ze, SEW and skewness

The set of input parameters in ANN #0 was chosen because they are similar to existing riming retrieval methods relying on MDV. In order to check whether adding more radar variables can improve the riming estimate, ANN #1 uses the SEW and skewness as additional input features. ANN #2 represents the set of input parameters if MDV can not be used e.g. due to

20   persistent up- or downdrafts.

In ~~Goodfellow et al. (2014), plots of accuracy vs. layer depth are used to determine the optimal ANN architecture. Following this concept, we plotted the root mean squared error (RMSE) vs. the number of neurons/ hidden layers (not shown) to find an architecture which is simple yet accurate. For the limited number of input features used in our approach, a relatively simple setup with one hidden layer and two neurons fulfills these requirements.In~~ the testing phase, the model performance is evaluated

25   using the test set prediction RMSE, in connection with visual inspection of scatter plots of ANN test set predictions and target values. In addition, time-height plots of ANN predictions are examined with regard to physical plausibility of the predicted features. We decided to put an additional focus on the ability of the networks to predict high FR values ~~> 0.5~~≥ 0.5. The reason for this choice is that ~~strong riming cases are less well represented in the training data set , but those cases are clearly~~ the extracted Doppler spectra features are expected to contain little information about riming up to a moderate riming stage.

30   For high FR values, the prediction should be more accurate because the riming signal should be clearer in the input features. Table 2 summarizes the test set RMSEs found for the three different ANN input parameter sets and the two radar frequencies. Fig. 2 shows scatter plots of ANN predictions and test set target values, for the three different ANN ensembles, for Ka- and W-band. The slopes of the linear models (red dashed lines in Fig. 2) are listed in Table 2. From Fig. 2 it becomes apparent
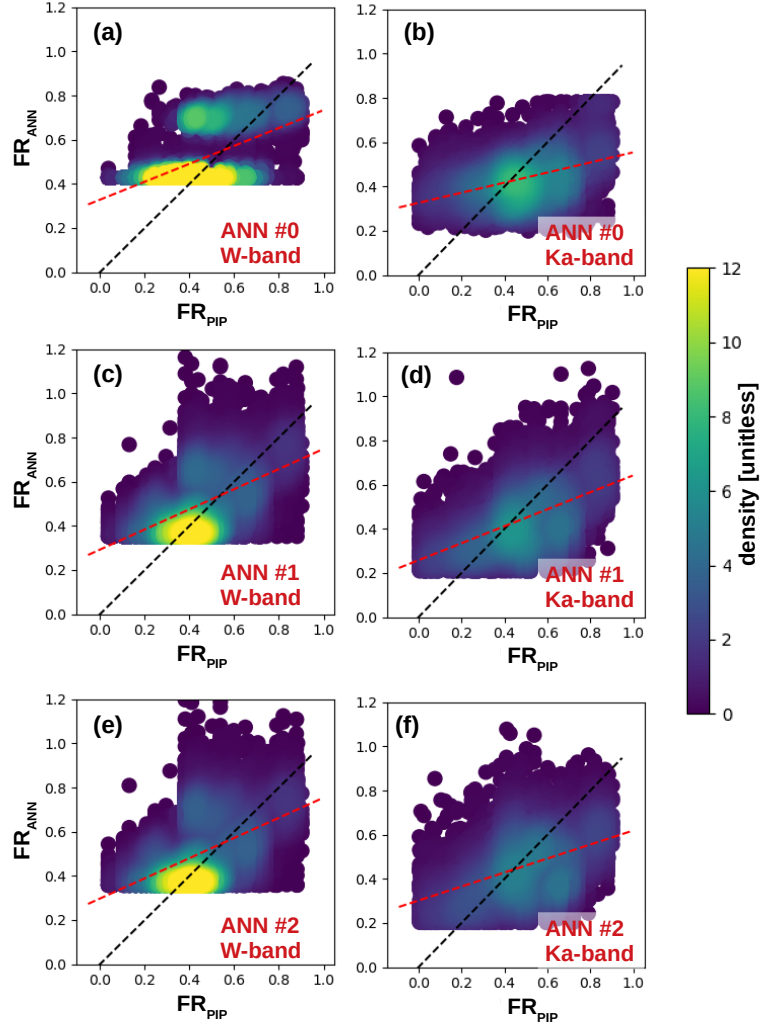
Figure 2. scatter plots of ANN performances on the test set. The black dashed line marks the 1:1 line, the red dashed line is a linear model fit to the data. The slopes of the red dashed lines are listed in Table 2. (a) W-band, ANN #0; (b) Ka-band, ANN #0 (c) W-band, ANN #1; (d) Ka-band, ANN #1; (e) W-band, ANN #2; (e) Ka-band, ANN #2

Table 2. Test set performances for the three different ANN ensembles, for the complete test set, and a subset of it containing only high (> 0.5) target values

ANN	test set RMSE Ka-band	test set RMSE Ka-band(FR > 0.5)	test set RMSE W-band	test set RMSE W-band(FR > 0.5)	slope Fig. 2 Ka-band	slope Fig. 2 W-band
ANN #0	~~0.22~~0.21	~~0.26~~0.23	0.15	0.15	0.23	0.41
ANN #1	0.18	~~0.16~~0.19	0.15	0.14	0.38	0.46
ANN #2	0.20	~~0.19~~0.20	0.15	0.15	0.32	0.46

that ANN #0 seems to have a problem in the W-band set up, because two populations of values are separated in the ~~input feature space (cf. Fig 1c).~~ predictions. The slopes of the linear fits, which should ideally be 1, are lowest for ANN #0 for both considered radar frequencies. This observation is however not reflected in the RMSEs listed in Table 2, which hardly differ between the three ANN ensembles for the W-band. The RMSEs in turn show that ANN #0 performs worse than the

5   other two setups for the Ka-band. This illustrates the limits of using a single quantity like RMSE as a quality metric. ANN #1 and #2 also have issues predicting high FR values, and all ANNs feature a cut-off at low FR values. This might be because riming signatures for low FR are not very pronounced in the cloud radar Doppler spectra, making it impossible for the ANNs to accurately predict $FR \leq 0.3$ from the extracted features. Similarly, the issues of the ANNs to predict high FR values could be explained by spectra of pure graupel not necessarily being bimodal, which would result in a less clear signature in SEW

10   and skewness. For the remaining study, we will focus only on ANN #1 and ANN #2. ~~Table 2 summarizes the test set RMSEs found for the three different ANN input parameter sets. The RMSE basically represents the error between the predicted and the retrieved FR. For ANN #0, the performance is much worse than for the other two setups, which is why, in the remaining study, we will focus only on ANN #1 and ANN #2. Due to the many steps involved in obtaining the predicted quantity, we will refer to it as a "riming index" rather than FR in the following.~~ We will refer to the predicted quantity in the following as FR_{ANN}, to

15   distinguish it from FR_{PIP}.

2.4   Error consideration

We acknowledge that several sources of uncertainty impact the riming predictions, only some of which can be quantified. For the training data set, assumptions about the density of unrimed snow and the viewing geometry corrected D_{max} were made. The m-D relation used to derive the mass of unrimed snow (Moisseev et al., 2017) is assumed to overestimate m_{us} by

20   a maximum of 5%. Furthermore, errors are introduced by the spatio-temporal matching of radar and PIP observations. To get a grasp of the overall effect of measurement uncertainties, which propagate into the FR_{ANN} predictions, a sensitivity study was performed. We assumed an error of 0.2 dBZ for Ze, one Doppler bin for MDV and SEW, and 0.2 for the skewness, and added this uncertainty to selected Doppler spectra. This was accomplished by random picking of n = 10,000 samples from a Gaussian distribution defined by the measured values as mean and above errors as standard deviation. The resulting standard

deviation in the predicted FR$_{ANN}$ is approximately 1%. In addition to the 5% uncertainty from the mass of unrimed snow, and other error sources which cannot be quantified we propose to assume an overall uncertainty of approximately 10% due to the training data. This uncertainty has to be added to the uncertainty of the ANN models (Table 2).

3   Results and Discussion

5   This section is structured as follows: In Section 3.1, a well-defined "benchmark" riming case from the BAECC dataset is presented. The application of the ANNs to the ~~TRIPEx-Pol~~ TRIPEx-pol dataset is demonstrated in Section 3.2. In Section 3.3, the performance of the ANNs for the W-band radar in Leipzig on 19 March 2021 is evaluated by comparison to in situ observations. Finally, in Section 3.4, we present ANN predictions for a case obtained during the DACAPO-PESO field campaign using the same W-band radar.

3.1   BAECC benchmark riming case

~~While~~ In addition to using the BAECC data set for training, validation and testing, we also capitalize on the fact that these data have already been extensively studied with respect to microphysical growth processes, and several well-defined case studies are available. We evaluate the performance of our newly developed riming estimation technique for a ~~1.5-hour~~ 1-hour period between 22:00 and 23:~~30~~15 UTC on 21 February ~~2014.~~ 2014, which is part of the 10% of the labeled data that were retained

15   for the test set. Riming was taking place starting at around ~~17.00~~17:00, the LWP reaching its maximum value of more than 1000 g m^{-2} around 22:00 UTC ~~(Fig. 4c, cf. Moisseev et al., 2017)~~(Fig. 4e, cf. Moisseev et al., 2017). Parts of the focus period considered here between 22:00 and 23:~~30~~15 UTC have been previously analyzed in detail e.g. by Kalesse et al. (2016, 2019), Moisseev et al. (2017), Mason et al. (2019) and Kneifel et al. (2015, 2016). Fig. 3 shows ~~the equivalent radar reflectivity (Ze)~~Ze, MDV, spectral width, skewness and SEW measured by the KAZR. The case is characterized by a "seeder-feeder" situation,

20   where a frontal snow cloud merges into a mid-level mixed-phase cloud. In the mixed-phase cloud, SLW layers are present at 0.7 to 0.9 km, and slightly below 3 km. As snow starts to fall from the frontal ("seeder") cloud into the lower-level ("feeder") cloud, intense riming happens along a slanted fall-streak feature at around 22:40 to 22:45 UTC, resulting in Doppler spectra with multiple peaks (Kalesse et al., 2016, 2019). During the following time period, a transition from strongly rimed particles to unrimed snow aggregates at between 23:03 and 23:10 UTC was observed by Kneifel et al. (2015, 2016); Moisseev et al.

25   (2017); Mason et al. (2019).

Fig. 4 shows the predicted ~~riming index~~ FR$_{ANN}$ for the two different ANN ensembles, ~~along with the measured liquid water path (LWP)~~and the two different radar frequencies, respectively. The ensembles yield very similar predictions, which is remarkable given the fact that ANN #2 does not use MDV as input feature. ~~Both approaches seem to predict too high values in the lowermost 500 m, where spectra are noisy and broadened by strong turbulence. Above this~~ The surface-induced turbulent

30   layer close to the ground ~~, which contains high values throughout, there are variations in the riming index: A~~ is masked with grey pixels, where EDR exceeds the threshold of 10^{-3} m^{2}s^{-3}. For the W-Band, columns where the estimated attenuation exceeded 10 dBZ, were masked. For all four configurations, a clear increase in ~~signal~~FR$_{ANN}$ is visible at around 22:40 UTC,
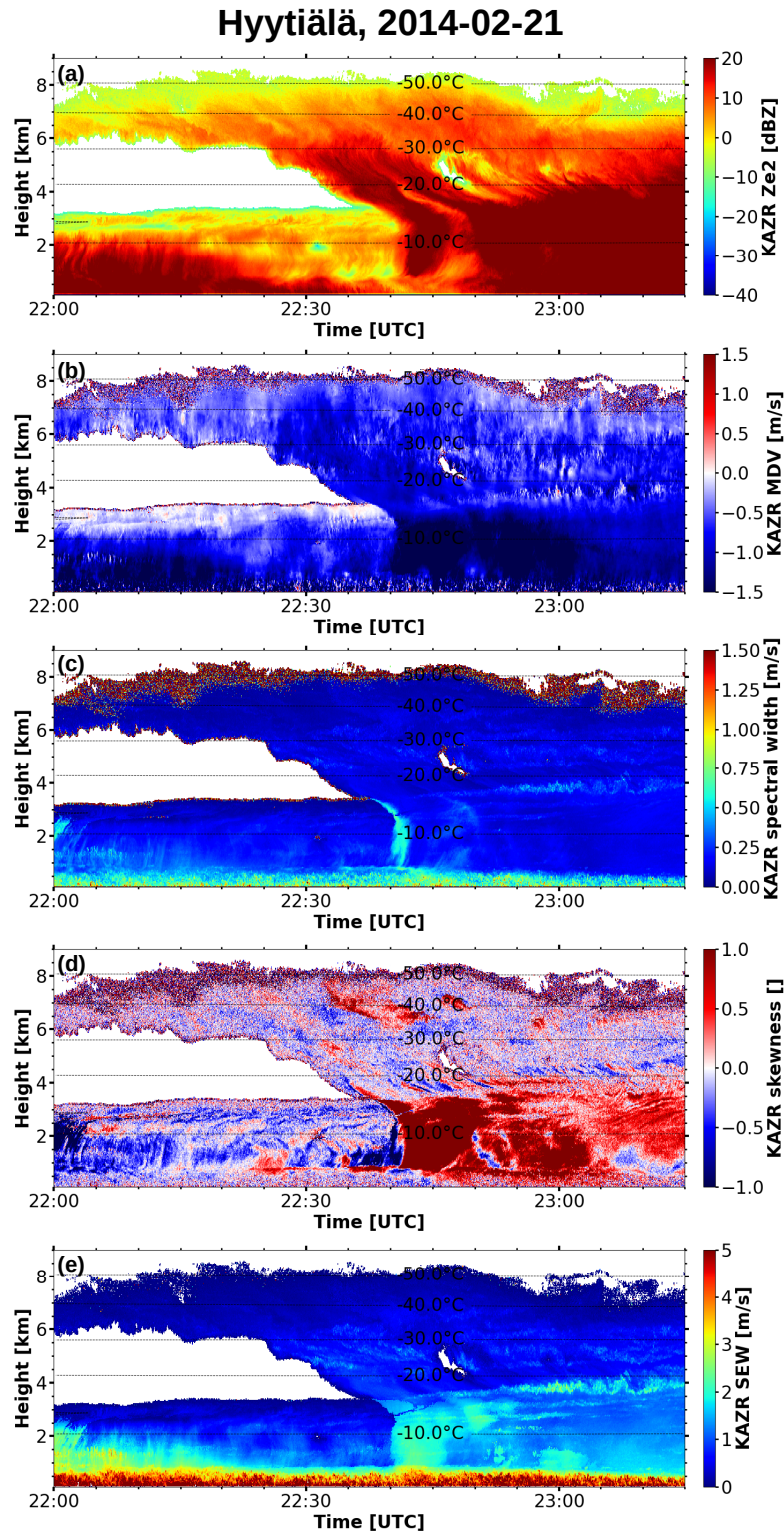
13

Figure 3. Radar moments measured by the KAZR in the focus period between 22:00 UTC and 23:~~30~~15 UTC on 21 February 2014. a) Equivalent radar reflectivity of the ~~main peak~~ full spectrum in dBZ; b) mean Doppler velocity (negative values indicating downward motion) computed from the full spectrum; c) spectral width computed from the full spectrum; d) skewness computed from the full spectrum; e) spectrum edge width.
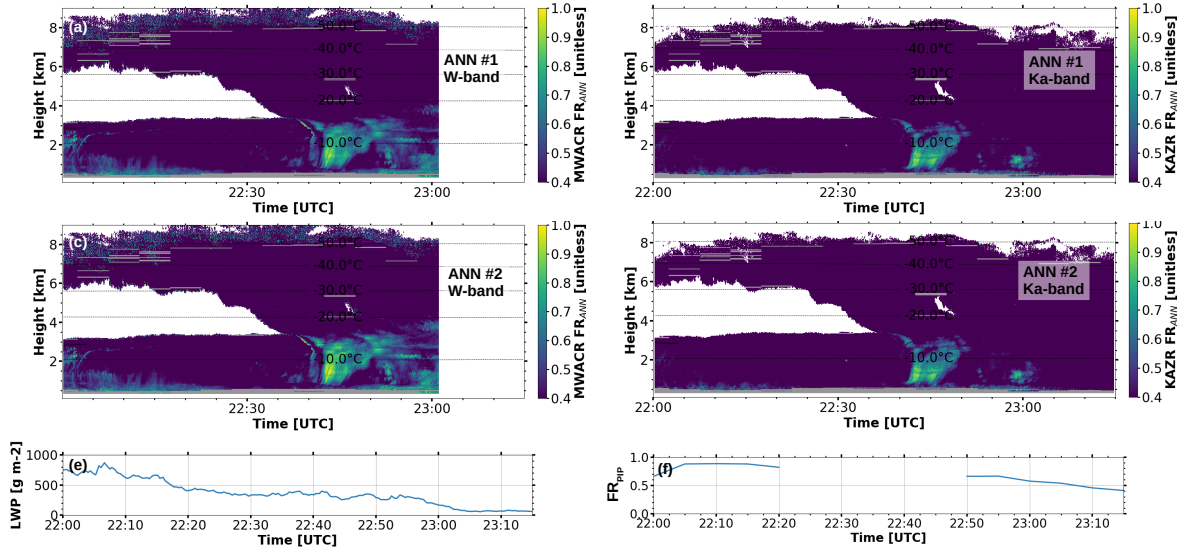
Figure 4. Riming during the focus case on 21 February 2014 predicted by (a) and (b) ANN # 1 trained on MWACR and KAZR data, respectively (Ze, MDV, SEW and skewness); (c) and (d) ANN # 2 trained on MWACR and KAZR data, respectively (Ze, SEW, and skewness); (e) Liquid water path obtained from microwave radiometer measurements; (f) FR_{PIP} measured during the event. Pixels with EDR above the threshold of $10^{-3} m^2 s^{-3}$ are masked in grey in (a)-(d).

when snow starts falling from the seeder cloud through the SLW layers in the lower-level mixed-phase (feeder) cloud. This is the period for which Kalesse et al. (2016, 2019) reported riming signatures in Doppler spectra featuring multiple peaks. Unfortunately, due to low precipitation intensity at the ground, no PIP-based FR retrieval was avaialble for this time period (Fig. 4f). However, a continuous decrease in LWP points to the depletion of SLW by the strong riming (Fig. 4c). Later, after around 22:50 UTC, the predictions differ between Ka- and W-band: While the two ANN ensembles trained on Ka-band data only predict riming during a short period around 23:00 UTC, the FR_{ANN} predictions in the W-band setup remain elevated up to a range of around 4 km. The Ka-band ANNs are apparently able to detect the transition from increased riming to less riming around 23:05 UTC, which was reported in existing studies of the event. Around this time, the LWP has reached its minimum, indicating that no SLW for further riming is available in the column. Coinciding with this period from around 23:05 UTC onwards, the FR_{ANN} is very low (≤ 0.5) throughout the cloud system, which is in accordance with the measured decrease of FR_{PIP} (Fig. 4f). This first case study is promising with respect to the usability of our trained ANNs to predict riming from cloud radar moments. Furthermore, the similarity between ANN#1 and ANN #2 for both radar frequencies shows that that their application might be possible even without the use of MDV.

3.2 ~~TRIPEx-Pol~~ TRIPEx-pol case study and triple-frequency signatures for seven cases

To answer the question whether the developed methods are able to generalize to conditions different than the ones they were trained on, considering another data set is required. For this reason, the two ANN sets are applied to data from a different site, obtained by radars with different settings ~~and a different (W-band) frequency~~. We will first focus in more detail on the

5 24 November 2018 case from the ~~TRIPEx-Pol~~ TRIPEx-pol campaign. This precipitation event has been analyzed with respect to rain and ice microphysics by Mróz et al. (2020), who found strong signatures of aggregation during the period from 06:45 to 07:45 UTC, and riming ~~after 08~~during a shorter time interval around 09:00 UTC.

Fig. 5 shows the radar moments measured during the period from ~~03~~02:00 UTC to ~~22:00~~11:30 UTC on 24 November 2018. The period characterized as "aggregation" by Mróz et al. (2020) clearly shows up as a patch of increased signal in the $DWR_{X,Ka}$

10 between 1 and 4 km range, while during the "riming" period around 09:00 UTC, an obvious increase in absolute MDV values can be observed in a similar range interval, along with an increase in SEW. In Fig. 6, the predictions of the two ANNs are shown side by side for W-band and Ka-band. In both cases, the CloudNet classification mask was used to only apply the ANNs to those parts of the cloud which were classified as ice or ice and liquid. For the two different frequencies, as well as for the two ANNs, the predicted ~~riming index is very similar and the features are almost identical~~FR_{ANN} features are very

15 similar. For the "aggregation" period, which has the strongest signal in the $DWR_{X,Ka}$ (Fig. 5c) from approximately 06:45 to 07:30 UTC, some riming is predicted, however relatively low values. During the "riming" period, which clearly shows up as increased MDV in Fig. 5b between 08:00 and 09:00 UTC, strong riming is predicted for both wavelengths, and by both ANN sets. Again, it is remarkable how the MDV features in Fig 5b can be discovered even in the predictions by ANN #2, which does not use MDV (two lower panels in Fig. 6). ~~These findings show that the ANNs can be applied at a different cloud~~

20 ~~radar frequency than the one they were trained at, indicating that the scattering properties of rimed particles are sufficiently similarbetween the two wavelengths considered. Apparently, sensitivity differences between the instruments do not play a major role for~~ Despite the differences in scattering properties and attenuation due to hydrometeors at Ka- and W-band as well as the different noise levels of the two radars, the ~~ANN performance, and the ANNs also do not seem to be impacted by the liquid precipitation below the cloud system. However, from Fig. 6, the possibility that in some cases aggregation is~~

25 ~~misclassified as riming cannot be outruled. To further investigate this question, we next take a look at statistics over seven cases observed during TRIPEx-Pol~~retrieved FR_{ANN} is rather similar. The common time-height grid in the TRIPEx-pol data set allows for convenient direct comparison of the two radar frequencies for each of the ANN set ups: The correlation of the predicted FR_{ANN} values for all considered TRIPEx-pol cases is high for both ANN #1 and #2, the R^2 being 0.73 and 0.81, respectively.

30 As mentioned earlier on, riming and aggregation can produce distinct signatures in the triple-frequency space ~~: While~~ of X-, Ka- and W-band reflectivity. When considering a plot with $DWR_{X,Ka}$ on the abscissa and $DWR_{Ka,W}$ on the ordinate, observations obtained during riming events fall onto a line at low $DWR_{X,Ka}$ ~~, aggregates~~ (pink line in Fig. 7). Aggregates, in contrast, tend to yield a hook-like feature. ~~These signatures are attributed to changes in the particle density during the riming process,~~ due to their comparably large size at which $DWR_{Ka,W}$ is in saturation because of Mie scattering. This conceptual
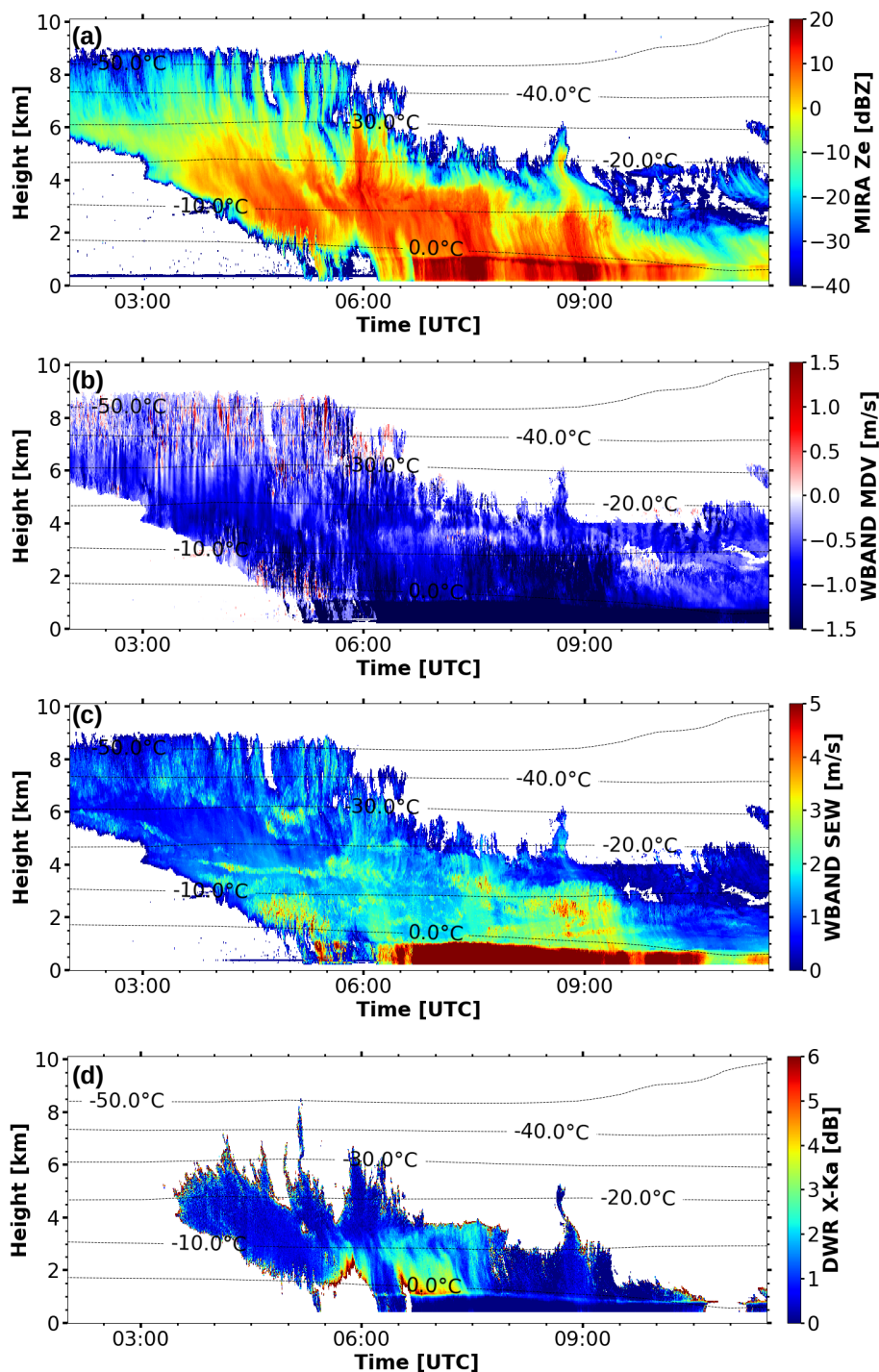
Figure 5. Radar moments measured during the 24 November 2018 case. a) Equivalent radar reflectivity measured by the Ka-band MIRA radar; b) mean Doppler velocity measured by the W-band radar; c) spectrum edge width derived from W-band Doppler spectra; d) Dual-Wavelengh-Ratio of X and Ka-band radar (DWR$_{X,Ka}$)
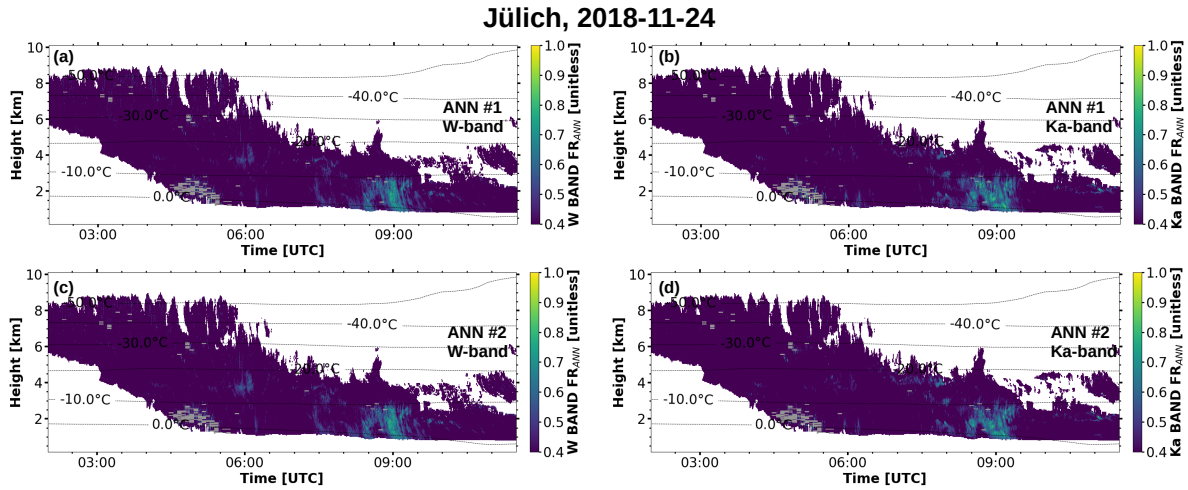
Figure 6. Predicted riming during the case on 24 November 2018 by ANN #1 and #2 using W-band moments (left column) and Ka-band moments (right column) as input features. The ANNs were applied only to those parts of the cloud which were classified as containing ice or ice and liquid by the CloudNet algorithm. Pixels with EDR above the threshold of $10^{-3} m^2 s^{-3}$ are masked in grey.

model was presented by Kneifel et al. (2015) and further explored by Mason et al. (2019), who found that not only the density, but also the shape parameter of the particle size distribution and the internal structure of aggregates can have strong impacts on triple-frequency signatures.

In Fig. 7, we plotted the observations for which an increased ~~riming index~~ $FR_{ANN} > 0.5$ was predicted by the ANNs on 2-D histograms in the triple-frequency space, colored by the median ~~riming index~~ FR_{ANN} of all the observations in the respective pixel. The pink line is drawn along the line of increasing median volume diameter expected for rimed particles according to Fig. 15 in Kneifel et al. (2015). With respect to frequency of occurrence (not shown here), the largest portion of the pixels for which a ~~riming index~~ $FR_{ANN} > 0.5$ was predicted, falls around that line, for both ANN sets and both frequencies. ~~For both ANN #1 and #2, there is even a trend towards higher riming indices with increasing $DWR_{Ka,W}$.~~ However, in all plots, there ~~is a population of points at higher $DWR_{X,Ka}$, which fall onto a hook-like feature, which is expected for aggregates in the triple-frequency space. The riming index for this population is smaller than for the pixels falling around the pink line, which is expected for riming.~~ This finding suggests that both ~~ANNs~~ ANN ensembles for both frequencies are capable of predicting ~~strong riming, but that there is uncertainty in distinguishing between aggregation and riming in other cases~~ elevated FR values.

3.3 Convective riming and aggregation case study in Leipzig

The findings in the previous sections motivate the need for additional comparisons to in situ observations. The 19 March 2021 Leipzig case is characterized by a wintertime convective mixed-phase cloud system, with cloud top at 3-4 km, and with embedded strong snow and graupel showers. Cold air aloft combined with some solar warming near the ground causes weak
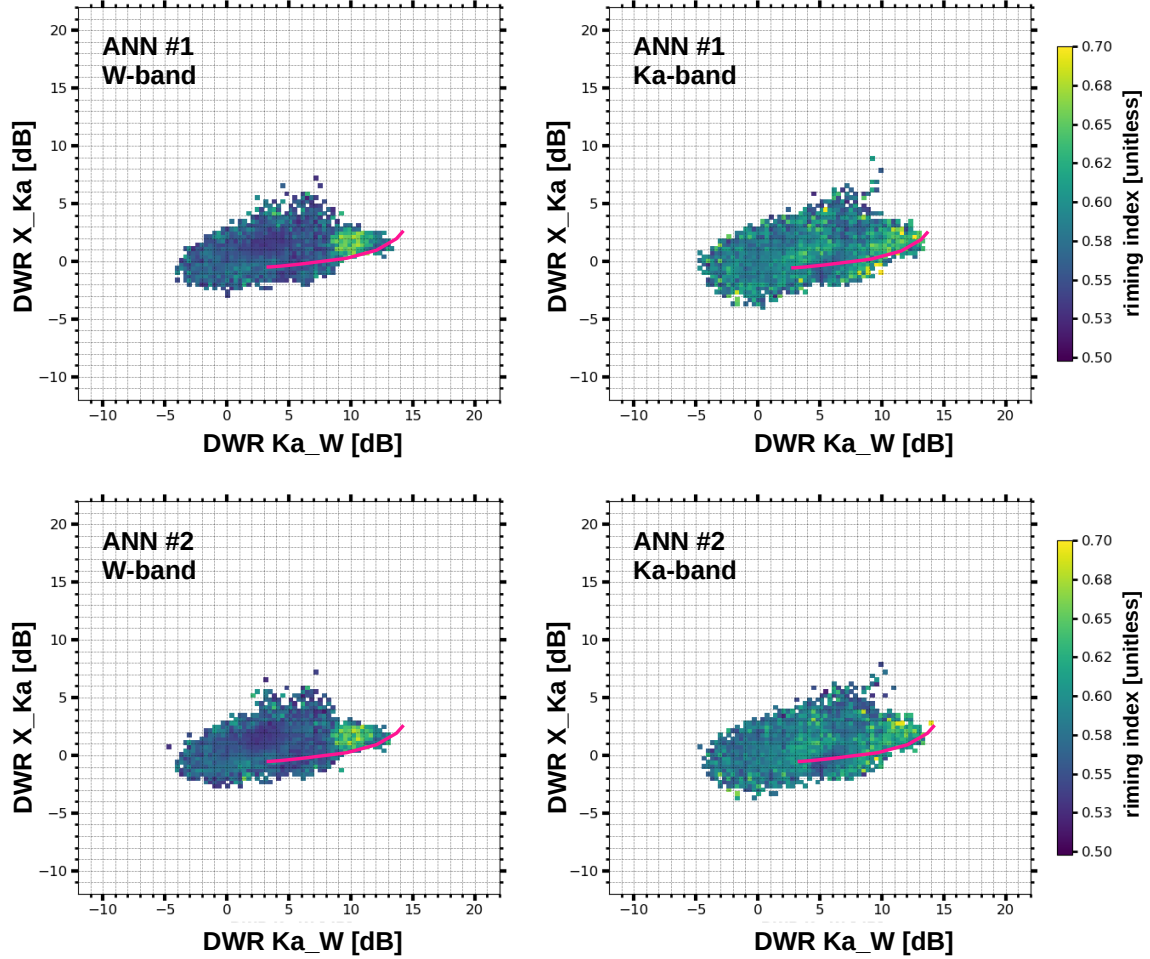
Figure 7. 2D histograms for seven cases chosen from the ~~TRIPEx-Pol~~ TRIPEx-pol data set where rime index > 0.5 was predicted. Each pixel contains at least 10 observations, and the color indicates the median ~~riming index~~ FR_{ANN} of all the observations contained in the pixel. ~~Riming index~~ FR_{ANN} predicted using W-band and Ka-band radar observations are displayed in the top and bottom panel plots, respectively. The pink line is drawn where rimed particles are expected according to Fig. 15 in Kneifel et al. (2015).
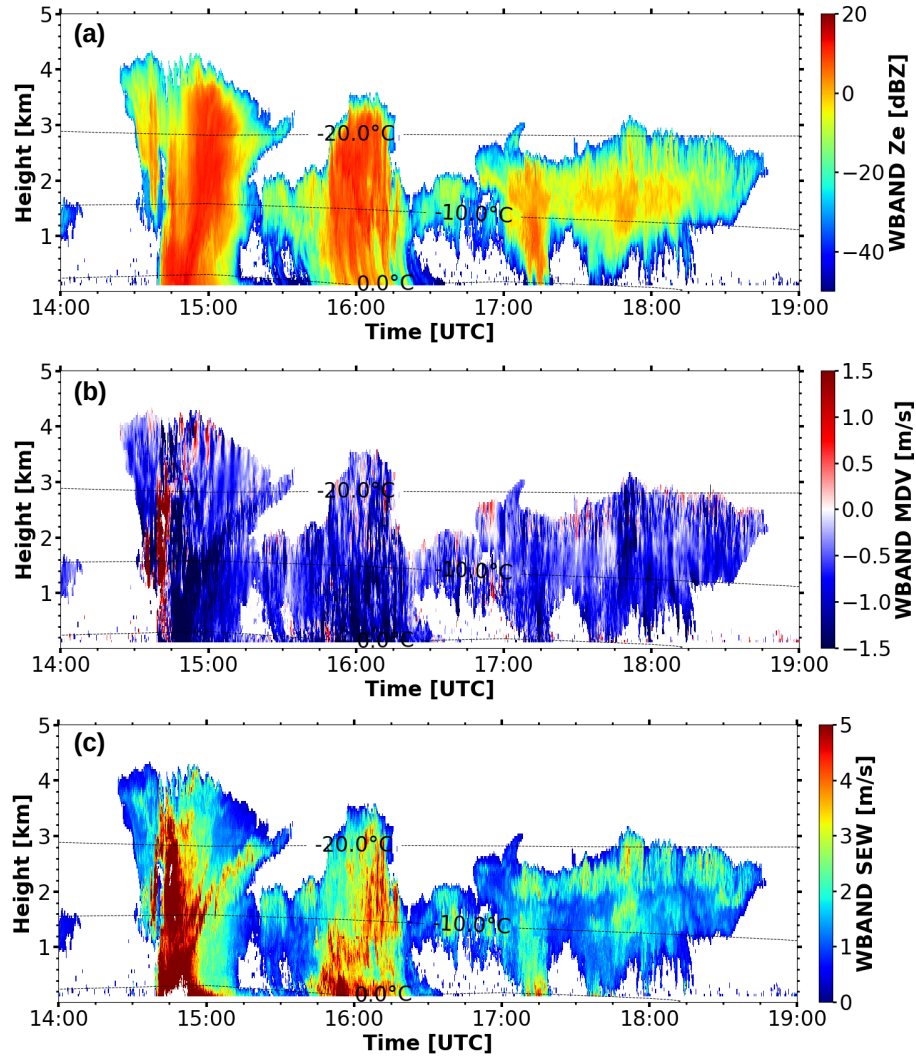
Figure 8. Radar moments measured by LIMRAD94 during the case on 19 March 2021. a) Equivalent radar reflectivity in dBZ; b) mean Doppler velocity; c) spectrum edge width

~~lability~~instability. Additional lifting, e.g. due to convergences, triggers showers and, as in this case study, is even sufficient to cause thunderstorms.

In Fig. 8, the first two moments of the cloud radar Doppler spectra during that day are shown, along with the SEW. Around 15:00 UTC, a strong updraft is visible in the MDV, followed by strongly negative Doppler velocities coinciding with increased
5 Ze and SEW values. Around that time, alternating graupel and snow showers were observed in the Leipzig area. This case is moderately convective, and most of the data would be excluded by filtering criteria such as the convection index κ used in the approach by Mosimann (1995).

Fig. 9 shows the ANN predictions for the time between 14:00 and 19:00 UTC. In the panels below, hydrometeors observed by the VISSS are shown for three selected time periods. In both ANN predictions, a clear transition from very high (around 1) to
10 very low (≤ 0.4) ~~riming index~~ FR_{ANN} values at around 15:00 UTC is visible. This change can be confirmed by the VISSS observations: In the period from 14:40 to 14:50 UTC, the images show dense and roundish particles (rimed aggregates, graupel). In contrast, during the period from 15:00 to 15:10 UTC, fluffy, unrimed aggregates are observed. Later, from approximately 15:45 UTC to 16:20 UTC, the ~~riming index~~ FR_{ANN} features increased values throughout the vertical column ($\approx 0.7 - 1.0$), but not as pronounced as in the period before 15:00 UTC. VISSS images taken during the period from 15:50 to 16:00 UTC reveal
15 a mixture of particles, which have different sizes and degrees of riming. Small hydrometeors with diameters well below 1 mm coincide with aggregates and graupel particles having sizes of several mm. The ANN predictions fit extremely well in line with these observations: In the first case, very high ~~riming index~~ FR_{ANN} values around 1 are predicted, whereas in the second case, the predicted riming indices are below 0.4. In the third case, ~~intermediate riming index~~ even though a portion of the column is masked due to the EDR threshold, it is visible that intermediate FR_{ANN} values are predicted by both ANNs. This leads to the
20 assumption that the ANNs are not only capable of detecting strong riming, but are also sensitive to the degree of riming, or the fraction of rimed particles compared to the total hydrometeor population. In this case, as well as in the previously presented results, the predictions of ANN #1 and #2 are strikingly similar. These findings show that predicting riming is possible even without the use of MDV. ~~Moreover, since~~ It has to be acknowledged, though, that the vertical distribution of FR_{ANN} cannot be verified by the VISSS observations. Since the ANNs proved to perform well for this wintertime convective case, this could
25 open a door to detect and even quantify riming in convective systems.

3.4 Punta Arenas gravity wave case

The previous findings make us confident that ANN #2 can be applied to the W-band radar data in the DACAPO-PESO data set. We do not apply ANN #1, because it would be biased by the orographic wave motions. Here, we analyze a case observed on 21 February 2019 (Fig. 10) from 13:30 to 22:00 UTC. A precipitating cloud system with cloud top around 2.5 to 3 km is present
30 from around 15 to 18 UTC. Above, a mid-level cloud with top around 6 km and varying cloud base is observed. Especially in the higher-level cloud, in the range between 3 and 6 km, a wave pattern is visible in the MDV (Fig. 10b), including a prominent downdraft around 18:00 UTC with MDV $\approx -2 \text{ ms}^{-1}$. Precipitation was reaching the ground between 15:00 and 18:00 UTC, and another precipitation event occurred around 21:00 UTC. At 16:30 UTC, marked by the red cross in Fig. 10d, graupel particles were observed at the ground on-site.
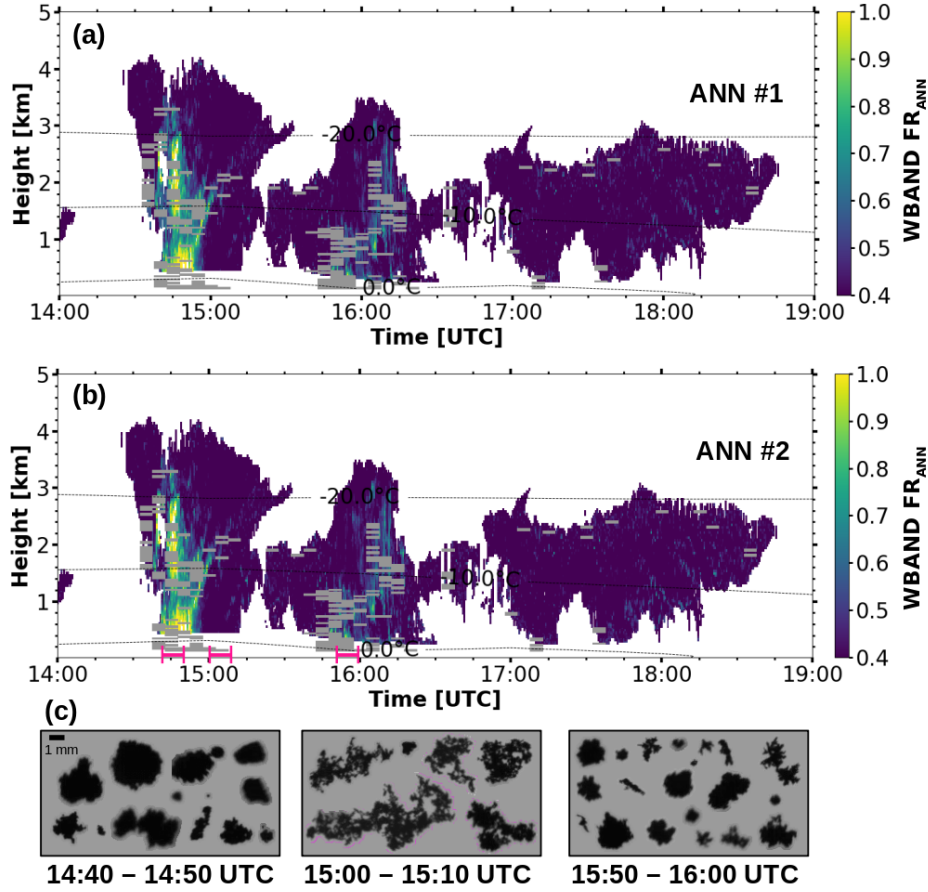
Figure 9. Riming during the case on 19 March 2021 predicted by (a) ANN #1 (SEW, Ze, skewness, MDV), (b) ANN #2 (SEW, Ze, skewness). The panels in (c) show images taken by the VISSS during the period from 14:40 to 14:50 UTC, the period from 15:00 to 15:10 UTC, and the period from 15:50 to 16:00 UTC. These periods are marked on the time axis in (b) with pink lines. Pixels with EDR above the threshold of $10^{-3} \text{m}^2 \text{s}^{-3}$ are masked in grey in (a) and (b).
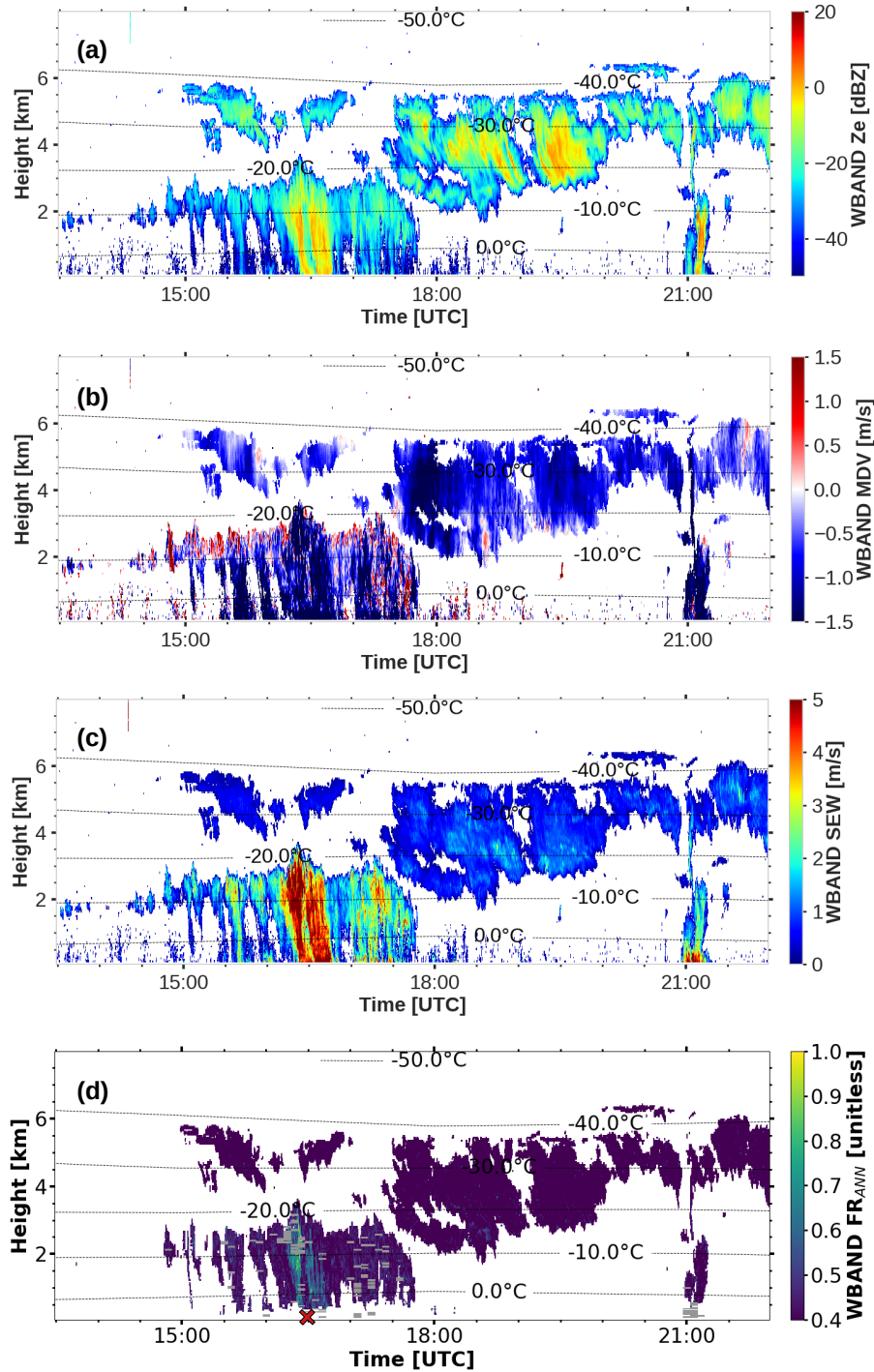
Figure 10. Radar moments measured by LIMRAD94 during the case on 21 February 2019. a) Equivalent radar reflectivity in dBZ; b) mean Doppler velocity; c) spectrum edge width and d) ~~riming index~~ FR_{ANN} predicted by ANN #2 (using, Ze, SEW and skewness). The red cross marks the time (16:30 UTC) when graupel particles were observed at the site.

In Fig. 10d, the ~~riming index~~ FR_{ANN} predicted by ANN #2 is shown. The predicted ~~riming index~~ FR_{ANN} is only increased in the lower part of the cloud system up to around 3.5 km range, which probably contains liquid water~~and reaches the maximum value of 1.0 around 16:~~. No increased FR_{ANN} values are predicted for the higher part of the cloud system. The maximum in FR_{ANN} predictions is reached around 16:30 ~~UTC. Later in the day, at around 21:00 to 21:15~~ UTC, ~~a fall streak reaching from the higher cloud system into the lower cloud with increased riming index is visible. This coincides with strongly negative values ≤ -3 m s^{-1} of measured MDV in Fig. 10b and precipitation at the surface.~~ coinciding with the observation of graupel particles.

4   Summary, Conclusions and Outlook

In this work, we have demonstrated the ability of artificial neural networks (ANNs) to estimate riming using ground-based zenith-pointing cloud radar measurements as input features. Training data were extracted from the BAECC data set by temporally matching PIP-based riming retrievals with cloud radar observations at cloud base. Ensembles of ANNs were trained to predict a ~~riming index~~FR_{ANN}, separately for Ka-band and W-band observations, using three different combinations of input variables: ANN #1 uses the equivalent radar reflectivity factor (Ze), the mean Doppler velocity (MDV), the spectrum edge width (SEW) and skewness as input features; ANN #2 uses Ze, SEW and skewness. One set of ANNs using only Ze and MDV as input features was not further considered due to low performance indicating that these two quantities are not sufficient for quantifying riming. We evaluated the trained models using four case studies and a longer data set comprising observations of seven mixed-phase cloud systems. In general, the predictions of ANN #1 and #2 were found to be very similar across all considered cases despite the different input variables. Both ANNs were able to predict strong riming and capture the subsequent transition to unrimed snow reported in literature for a case from the BAECC experiment (Fig. 4). ~~One limitation was identified within the surface-induced turbulent layer, where Doppler spectra are broadened and the ANN estimates of the riming index are too high~~ A turbulence threshold of EDR = 10^{-3} m^{2}s^{-3} was applied to prevent too high FR_{ANN} ANN estimates due to broadened Doppler spectra. It was shown that the models are able to generalize to a new data set, i.e. different radar systems ~~including a different frequency~~ for both considered wavelengths (Fig. 6~~, W-band) than the one they were trained on (Ka-band~~), and different meteorological conditions. ANN predictions for seven cloud cases were shown to match expected signatures of riming in the triple-frequency observation space of X-, Ka- and W-band (Fig. 7). Large ~~riming index~~ FR_{ANN} values mostly fall into the region for which riming is expected~~: The retrieved values are increasing along the line of increasing median volume diameter reported for rimed particles. Predictions falling into the region where aggregation is expected are lower in magnitude, and occur less frequent~~. The application of both ANNs to a convective wintertime cloud case showed that the method can also be applied to convective systems (Fig. 9). Because ANN #2 does not depend on MDV, it was applied to an orographic case, yielding high ~~riming index~~ FR_{ANN} values for the period during which solid graupel particles were observed at the site (Fig. 10). These findings indicate that retrieving riming is possible even without the use of MDV.

~~This study closes an important gap in our abilities to quantify the riming process with cloud radars. Future work will focus on the application of the newly developed technique to longer-term data sets to investigate the drivers of riming, including~~

~~orographic conditions.~~ One of the major constraints of this study is the limited training data set. As better and longer-term training data sets become available, the ML techniques at hand can be more fully exploited and further improvement of the ANN performance is expected. Also, ~~defining the turbulence threshold that prohibits the application of the method to convective systems will be a subject of future studies~~ an extended data set would better allow to quantify the errors of the ML method and to understand the limitations with respect to identifying very high and very low FR values.

This study closes an important gap in our abilities to quantify the riming process with cloud radars. Further validation, e.g. by comparison of this technique with airborne in situ observations would be a useful extension of this work. Future applications will focus on longer-term data sets to investigate the drivers of riming, including orographic conditions.

References

Barrett, A. I., Westbrook, C. D., Nicol, J. C., and Stein, T. H. M.: Rapid ice aggregation process revealed through triple-wavelength Doppler spectrum radar analysis, Atmos. Chem. Phys., 19, 5753–5769, https://doi.org/10.5194/acp-19-5753-2019, 2019.

Bühl, J., Seifert, P., Myagkov, A., and Ansmann, A.: Measuring ice- and liquid-water properties in mixed-phase cloud layers at the Leipzig Cloudnet station, Atmos. Chem. Phys., 16, 10 609–10 620, https://doi.org/10.5194/acp-16-10609-2016, 2016.

Cao, Y., Tan, W., and Wu, Z.: Aircraft icing: An ongoing threat to aviation safety, Aerosp. Sci. Technol., 75, 353–385, https://doi.org/10.1016/j.ast.2017.12.028, 2018.

Dias Neto, J., Kneifel, S., Ori, D., Trömel, S., Handwerker, J., Bohn, B., Hermes, N., Mühlbauer, K., Lenefer, M., and Simmer, C.: The TRIple-frequency and Polarimetric radar Experiment for improving process observation of winter precipitation, Earth Syst. Sci. Data, 11, 845–863, https://doi.org/10.5194/essd-11-845-2019, 2019.

Floutsi, A. A., Baars, H., Radenz, M., Haarig, M., Yin, Z., Seifert, P., Jimenez, C., Ansmann, A., Engelmann, R., Barja, B., Zamorano, F., and Wandinger, U.: Advection of Biomass Burning Aerosols towards the Southern Hemispheric Mid-Latitude Station of Punta Arenas as Observed with Multiwavelength Polarization Raman Lidar, Remote Sensing, 13, https://doi.org/10.3390/rs13010138, 2021.

Foth, A., Kanitz, T., Engelmann, R., Baars, H., Radenz, M., Seifert, P., Barja, B., Fromm, M., Kalesse, H., and Ansmann, A.: Vertical aerosol distribution in the southern hemispheric midlatitudes as observed with lidar in Punta Arenas, Chile ($53.2°$ S and $70.9°$ W), during ALPACA, Atmos. Chem. Phys., 19, 6217–6233, https://doi.org/10.5194/acp-19-6217-2019, 2019.

Garrett, T. J., Yuter, S. E., Fallgatter, C., Shkurko, K., Rhodes, S. R., and Endries, J. L.: Orientations and aspect ratios of falling snow, Geophys. Res. Lett., 42, 4617–4622, https://doi.org/https://doi.org/10.1002/2015GL064040, 2015.

Goodfellow, I., Bengio, Y., and Courville, A.: Deep Learning, MIT Press, 2016.

Goodfellow, I. J., Bulatov, Y., Ibarz, J., Arnoud, S., and Shet, V.: Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks, 2014.

Görsdorf, U., Lehmann, V., Bauer-Pfundstein, M., Peters, G., Vavriv, D., Vinogradov, V., and Volkov, V.: A 35-GHz Polarimetric Doppler Radar for Long-Term Observations of Cloud Parameters—Description of System and Data Processing, J. Atmos. Oceanic Technol., 32, 675–690, https://doi.org/10.1175/jtech-d-14-00066.1, 2015.

Heymsfield, A. J.: A Comparative Study of the Rates of Development of Potential Graupel and Hail Embryos in High Plains Storms, Journal of Atmospheric Sciences, 39, 2867 – 2897, https://doi.org/10.1175/1520-0469(1982)039<2867:ACSOTR>2.0.CO;2, 1982.

Heymsfield, A. J., Schmitt, C., Chen, C.-C.-J., Bansemer, A., Gettelman, A., Field, P. R., and Liu, C.: Contributions of the Liquid and Ice Phases to Global Surface Precipitation: Observations and Global Climate Modeling, J. Atmos. Sci., 77, 2629 – 2648, https://doi.org/10.1175/JAS-D-19-0352.1, 2020.

Hildebrand, P. H. and Sekhon, R. S.: Objective Determination of the Noise Level in Doppler Spectra, J. Appl. Meteor., 13, 808–811, https://doi.org/10.1175/1520-0450(1974)013<0808:odotnl>2.0.co;2, 1974.

Houze, R. A. and Medina, S.: Turbulence as a Mechanism for Orographic Precipitation Enhancement, J. Atmos. Sci., 62, 3599 – 3623, https://doi.org/10.1175/JAS3555.1, 2005.

Illingworth, A. J., Hogan, R. J., O'Connor, E., Bouniol, D., Brooks, M. E., Delanoé, J., Donovan, D. P., Eastment, J. D., Gaussiat, N., Goddard, J. W. F., Haeffelin, M., Baltink, H. K., Krasnov, O. A., Pelon, J., Piriou, J.-M., Protat, A., Russchenberg, H. W. J., Seifert, A., Tompkins, A. M., van Zadelhoff, G.-J., Vinit, F., Willén, U., Wilson, D. R., and Wrench, C. L.: Cloudnet: Continuous Evaluation of Cloud Profiles in

Seven Operational Models Using Ground-Based Observations, B. Am. Meteorol. Soc., 88, 883–898, https://doi.org/10.1175/BAMS-88-6-883, 2007.

Kalesse, H., Szyrmer, W., Kneifel, S., Kollias, P., and Luke, E.: Fingerprints of a riming event on cloud radar Doppler spectra: observations and modeling, Atmos. Chem. Phys., 16, 2997–3012, https://doi.org/10.5194/acp-16-2997-2016, 2016.

5   Kalesse, H., Vogl, T., Paduraru, C., and Luke, E.: Development and validation of a supervised machine learning radar Doppler spectra peak finding algorithm, Atmos. Meas. Tech., https://doi.org/10.5194/amt-2019-48, 2019.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, 2017.

Kneifel, S. and Moisseev, D.: Long-Term Statistics of Riming in Nonconvective Clouds Derived from Ground-Based Doppler Cloud Radar Observations, J. Atmos. Sci., 77, 3495–3508, https://doi.org/10.1175/JAS-D-20-0007.1, 2020.

10   Kneifel, S., von Lerber, A., Tiira, J., Moisseev, D., Kollias, P., and Leinonen, J.: Observed relations between snowfall microphysics and triple-frequency radar measurements, J. Geophys. Res. Atmos., 120, 6034–6055, https://doi.org/10.1002/2015jd023156, 2015.

Kneifel, S., Kollias, P., Battaglia, A., Leinonen, J., Maahn, M., Kalesse, H., and Tridon, F.: First observations of triple-frequency radar Doppler spectra in snowfall: Interpretation and applications, Geophys. Res. Lett., 43, 2225–2233, https://doi.org/10.1002/2015gl067618, 2016.

15   Kollias, P., Miller, M. A., Luke, E. P., Johnson, K. L., Clothiaux, E. E., Moran, K. P., Widener, K. B., and Albrecht, B. A.: The Atmospheric Radiation Measurement Program Cloud Profiling Radars: Second-Generation Sampling Strategies, Processing, and Cloud Data Products, J. Atmos. Ocean. Tech., 24, 1199–1214, https://doi.org/10.1175/jtech2033.1, 2007.

Küchler, N., Kneifel, S., Löhnert, U., Kollias, P., Czekala, H., and Rose, T.: A W-Band Radar–Radiometer System for Accurate and Continuous Monitoring of Clouds and Precipitation, J. Atmos. Oceanic Technol., 34, 2375–2392, https://doi.org/10.1175/jtech-d-17-0019.1, 20   2017.

Lamb, D. and Verlinde, J.: Physics and Chemistry of Clouds, Cambridge University Press, https://doi.org/10.1017/CBO9780511976377, 2011.

Leinonen, J. and Szyrmer, W.: Radar signatures of snowflake riming: A modeling study, Earth and Space Science, 2, 346–358, https://doi.org/https://doi.org/10.1002/2015EA000102, 2015.

25   Li, H., Moisseev, D., and von Lerber, A.: How Does Riming Affect Dual-Polarization Radar Observations and Snowflake Shape?, J. Geophys. Res. Atmos., 123, 6070–6081, https://doi.org/https://doi.org/10.1029/2017JD028186, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017JD028186, 2018.

Li, H., Tiira, J., von Lerber, A., and Moisseev, D.: Towards the connection between snow microphysics and melting layer: insights from multifrequency and dual-polarization radar observations during BAECC, Atmos. Chem. Phys., 20, 9547–9562, https://doi.org/10.5194/acp-30   20-9547-2020, 2020.

Luke, E. P., Kollias, P., Johnson, K. L., and Clothiaux, E. E.: A Technique for the Automatic Detection of Insect Clutter in Cloud Radar Returns, J. Atmos. Ocean. Tech., 25, 1498–1513, https://doi.org/10.1175/2007JTECHA953.1, 2008.

Luke, E. P., Kollias, P., and Shupe, M. D.: Detection of supercooled liquid in mixed-phase clouds using radar Doppler spectra, J. Geophys. Res. Atmos., 115, https://doi.org/10.1029/2009JD012884, 2010.

35   Löhnert, U., Schween, J. H., Acquistapace, C., Ebell, K., Maahn, M., Barrera-Verdejo, M., Hirsikko, A., Bohn, B., Knaps, A., O'Connor, E., Simmer, C., Wahner, A., and Crewell, S.: JOYCE: Jülich Observatory for Cloud Evolution, Bull. Amer. Meteor. Soc., 96, 1157 – 1174, https://doi.org/10.1175/BAMS-D-14-00105.1, 2015.

Maahn, M. and Löhnert, U.: Potential of Higher-Order Moments and Slopes of the Radar Doppler Spectrum for Retrieving Microphysical and Kinematic Properties of Arctic Ice Clouds, J. Appl. Meteor. Climatol., 56, 263–282, https://doi.org/10.1175/JAMC-D-16-0020.1, 2017.

Maahn, M., Löhnert, U., Kollias, P., Jackson, R. C., and McFarquhar, G. M.: Developing and Evaluating Ice Cloud Parameterizations for Forward Modeling of Radar Moments Using in situ Aircraft Observations, J. Atmos. Oceanic Technol., 32, 880–903, https://doi.org/10.1175/JTECH-D-14-00112.1, 2015.

Maahn, M., Radenz, M., Cox, C., Gallagher, M., Hutchings, J., Shupe, M., and Uttal, T.: Measuring Snowfall Properties with the Video In Situ Snowfall Sensor during MOSAiC, EGU21 Abstr., https://doi.org/10.5194/egusphere-egu21-3306, 2021.

Mason, S. L., Chiu, C. J., Hogan, R. J., Moisseev, D., and Kneifel, S.: Retrievals of Riming and Snow Density From Vertically Pointing Doppler Radars, J. Geophys. Res. Atmos., https://doi.org/10.1029/2018jd028603, 2018.

Mason, S. L., Hogan, R. J., Westbrook, C. D., Kneifel, S., Moisseev, D., and von Terzi, L.: The importance of particle size distribution and internal structure for triple-frequency radar retrievals of the morphology of snow, Atmos. Meas. Tech., 12, 4993–5018, https://doi.org/10.5194/amt-12-4993-2019, https://amt.copernicus.org/articles/12/4993/2019/, 2019.

Matrosov, S.: Assessment of Radar Signal Attenuation Caused by the Melting Hydrometeor Layer, IEEE Trans. Geosci. Remote Sens., 46, 1039–1047, https://doi.org/10.1109/tgrs.2008.915757, 2008.

Mech, M., Maahn, M., Kneifel, S., Ori, D., Orlandi, E., Kollias, P., Schemann, V., and Crewell, S.: PAMTRA 1.0: The Passive and Active Microwave Radiative TRAnsfer Tool for Simulating Radiometer and Radar Measurements of the Cloudy Atmosphere, Geosci. Model Dev., 13, 4229–4251, https://doi.org/10.5194/gmd-13-4229-2020, 2020.

Moisseev, D.: Snowfalke mass retrievals 2014/2015, https://github.com/dmoisseev/Snow-Retrievals-2014-2015/tree/v1.0, https://github.com/dmoisseev/Snow-Retrievals-2014-2015/tree/v1.0, 2018.

Moisseev, D., von Lerber, A., and Tiira, J.: Quantifying the effect of riming on snowfall using ground-based observations, J. Geophys. Res. Atmos., 122, 4019–4037, https://doi.org/10.1002/2016JD026272, 2017.

Morrison, H. and Milbrandt, J. A.: Parameterization of Cloud Microphysics Based on the Prediction of Bulk Ice Particle Properties. Part I: Scheme Description and Idealized Tests, J. Atmos. Sci., 72, 287–311, https://doi.org/10.1175/jas-d-14-0065.1, 2015.

Morrison, H., van Lier-Walqui, M., Fridlind, A. M., Grabowski, W. W., Harrington, J. Y., Hoose, C., Korolev, A., Kumjian, M. R., Milbrandt, J. A., Pawlowska, H., Posselt, D. J., Prat, O. P., Reimel, K. J., Shima, S.-I., van Diedenhoven, B., and Xue, L.: Confronting the Challenge of Modeling Cloud and Precipitation Microphysics, J. Adv. Model. Earth Syst., 12, e2019MS001 689, https://doi.org/10.1029/2019MS001689, 2020.

Mosimann, L.: An improved method for determining the degree of snow crystal riming by vertical Doppler radar, Atmos. Res., 37, 305–323, https://doi.org/https://doi.org/10.1016/0169-8095(94)00050-N, 1995.

Mosimann, L., Weingartner, E., and Waldvogel, A.: An analysis of accreted drop sizes and mass on rimed snow crystals., 1994.

Mróz, K., Battaglia, A., Kneifel, S., D'Adderio, L. P., and Dias Neto, J.: Triple-Frequency Doppler Retrieval of Characteristic Raindrop Size, Earth Space Sci., 7, e2019EA000 789, https://doi.org/10.1029/2019EA000789, 2020.

Mülmenstädt, J., Sourdeval, O., Delanoë, J., and Quaas, J.: Frequency of occurrence of rain from liquid-, mixed-, and ice-phase clouds derived from A-Train satellite retrievals, Geophys. Res. Lett., 42, 6502–6509, https://doi.org/10.1002/2015GL064604, 2015.

Oue, M., Kollias, P., Matrosov, S. Y., Battaglia, A., and Ryzhkov, A. V.: Combination Analysis of Multi-Wavelength, Multi-Parameter Radar Measurements for Snowfall, Atmos. Meas. Tech. Discuss., 2021, 1–33, https://doi.org/10.5194/amt-2021-78, 2021.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res., 12, 2825–2830, 2011.

Petäjä, T., O'Connor, E. J., Moisseev, D., Sinclair, V. A., Manninen, A. J., Väänänen, R., von Lerber, A., Thornton, J. A., Nicoll, K., Petersen, W., Chandrasekar, V., Smith, J. N., Winkler, P. M., Krüger, O., Hakola, H., Timonen, H., Brus, D., Laurila, T., Asmi, E., Riekkola, M.-L., Mona, L., Massoli, P., Engelmann, R., Komppula, M., Wang, J., Kuang, C., Bäck, J., Virtanen, A., Levula, J., Ritsche, M., and Hickmon, N.: BAECC: A Field Campaign to Elucidate the Impact of Biogenic Aerosols on Clouds and Climate, B. Am. Meteorol. Soc., 97, 1909–1928, https://doi.org/10.1175/bams-d-14-00199.1, 2016.

Pettersen, C., Bliven, L. F., von Lerber, A., Wood, N. B., Kulie, M. S., Mateling, M. E., Moisseev, D. N., Munchak, S. J., Petersen, W. A., and Wolff, D. B.: The Precipitation Imaging Package: Assessment of Microphysical and Bulk Characteristics of Snow, Atmosphere, 11, https://doi.org/10.3390/atmos11080785, 2020.

Protat, A., Rauniyar, S., Delanoë, J., Fontaine, E., and Schwarzenboeck, A.: W-Band (95 GHz) Radar Attenuation in Tropical Stratiform Ice Anvils, J. Atmos. Oceanic Technol., 36, 1463 – 1476, https://doi.org/10.1175/JTECH-D-18-0154.1, 2019.

Radenz, M., Bühl, J., Seifert, P., Baars, H., Engelmann, R., González, B. B., Mamouri, R.-E., Zamorano, F., and Ansmann, A.: Hemispheric contrasts in ice formation in stratiform mixed-phase clouds: Disentangling the role of aerosol and dynamics with ground-based remote sensing, submitted to Atmos. Chem. Phys. Discuss., 2021.

Seifert, A. and Rasp, S.: Potential and Limitations of Machine Learning for Modeling Warm-Rain Cloud Microphysical Processes, J. Adv. Model. Earth Syst., 12, e2020MS002 301, https://doi.org/https://doi.org/10.1029/2020MS002301, 2020.

Seifert, A., Leinonen, J., Siewert, C., and Kneifel, S.: The Geometry of Rimed Aggregate Snowflakes: A Modeling Study, J. Adv. Model. Earth Syst., 0, https://doi.org/10.1029/2018MS001519, 2019.

Serke, D. J., Reehorst, A. L., and Politovich, M. K.: Supercooled large drop detection with NASA's Icing Remote Sensing System, in: Remote Sensing of Clouds and the Atmosphere XV, edited by Picard, R. H., Schäfer, K., Comeron, A., and van Weele, M., SPIE, https://doi.org/10.1117/12.863176, 2010.

Tan, I., Storelvmo, T., and Zelinka, M. D.: Observational constraints on mixed-phase clouds imply higher climate sensitivity, Science, 352, 224–227, https://doi.org/10.1126/science.aad5300, 2016.

Torgo, L., Ribeiro, R. P., Pfahringer, B., and Branco, P.: SMOTE for Regression, in: Progress in Artificial Intelligence, pp. 378–389, Springer Berlin Heidelberg, https://doi.org/10.1007/978-3-642-40669-0_33, 2013.

Tukiainen, S., O'Connor, E., and Korpinen, A.: CloudnetPy: A Python package for processing cloud remote sensing data, J. Open Source Softw., 5, 2123, https://doi.org/10.21105/joss.02123, 2020.

van den Heuvel, F., Foresti, L., Gabella, M., Germann, U., and Berne, A.: Learning about the vertical structure of radar reflectivity using hydrometeor classes and neural networks in the Swiss Alps, Atmos. Meas. Tech., 13, 2481–2500, https://doi.org/10.5194/amt-13-2481-2020, 2020.

Vogel, J. M. and Fabry, F.: Contrasting Polarimetric Observations of Stratiform Riming and Nonriming Events, J. Appl. Meteorol. Clim., 57, 457–476, https://doi.org/10.1175/JAMC-D-16-0370.1, 2018.

von Lerber, A., Moisseev, D., Bliven, L. F., Petersen, W., Harri, A.-M., and Chandrasekar, V.: Microphysical Properties of Snow and Their Link to Ze–S Relations during BAECC 2014, J. Appl. Meteorol. Clim., 56, 1561 – 1582, https://doi.org/10.1175/JAMC-D-16-0379.1, 2017.

Woods, C. P., Stoelinga, M. T., Locatelli, J. D., and Hobbs, P. V.: Microphysical Processes and Synergistic Interaction between Frontal and Orographic Forcing of Precipitation during the 13 December 2001 IMPROVE-2 Event over the Oregon Cascades, J. Atmos. Sci., 62, 3493 – 3519, https://doi.org/10.1175/JAS3550.1, 2005.