Dual-frequency spectral radar retrieval of snowfall microphysics: a physically constrained deep learning approach

Anne-Claire Billault-Roux[1], Gionata Ghiggi[1], Louis Jaffeux[2], Audrey Martini[3], Nicolas Viltard[3], and Alexis Berne[1]

[1]Environmental Remote Sensing Laboratory, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[2]Laboratoire de Météorologie Physique, Aubière, France
[3]Laboratoire Atmosphère, Milieux et Observations Spatiales, IPSL, UVSQ Université Paris-Saclay, Sorbonne Université, CNRS, Guyancourt, France

Correspondence: alexis.berne@epfl.ch

Abstract. The use of meteorological radars to study snowfall microphysical properties and processes is well established, in particular through two techniques: the use of multi-frequency radar measurements and the analysis of radar Doppler spectra. We propose a novel approach to retrieve snowfall properties by combining both techniques, while relaxing some assumptions on e.g. beam matching and non-turbulent atmosphere.

5   The method relies on a two-step deep-learning framework inspired from data compression techniques: an encoder model maps a high-dimensional signal to a lower-dimensional "latent" space, while the decoder reconstructs the original signal from this latent space. Here, Doppler spectrograms at two frequencies constitute the high-dimensional input, while the latent features are constrained to represent the snowfall properties of interest. The decoder network is first trained to emulate Doppler spectra from a set of microphysical variables, using simulations from the radiative transfer model PAMTRA as training data. In a second

10   step, the encoder network learns the inverse mapping, from real measured dual-frequency spectrograms to the microphysical latent space; doing so, it leverages the spatial consistency of the measurements to mitigate the problem's ill-posedness.

The method was implemented on X- and W-band data from the ICE GENESIS campaign that took place in the Swiss Jura in January 2021. An in-depth assessment of the retrieval's accuracy was performed through comparisons with colocated aircraft in-situ measurements collected during 3 precipitation events. The agreement is overall good and opens up possibilities for acute

15   characterization of snowfall microphysics on larger datasets. A discussion of the method's sensitivity and limitations is also conducted.

The main contribution of this work is on the one hand the theoretical framework itself, which can be applied to other remote sensing retrieval applications and is thus possibly of interest to a broad audience across atmospheric sciences. On the other hand, the retrieved seven microphysical descriptors provide relevant insights into snowfall processes.

1   Introduction

Solid precipitation is a phenomenon of extraordinary complexity, whose better understanding and modeling remains a key challenge in atmospheric science. A more accurate representation of snowfall microphysical processes is not only crucial to

improve weather forecast models and precipitation quantification (e.g. Khain et al., 2015; Morrison et al., 2020), but also to reduce current uncertainties in cloud radiative properties, with in turn sizeable impacts on climate-oriented research (e.g. Curry

25 et al., 1996; Matus and L'Ecuyer, 2017). From a different perspective, snowfall microphysics is also relevant to a wide range of socio-economical fields, including the aviation industry, for which ensuring flight safety in snowfall conditions is critical (Rasmussen et al., 2000; Cao et al., 2018; Taszarek et al., 2020).

Unfortunately, the quantification of snowfall properties, such as particle size, mass, bulk density and geometry, is not a straightforward task. In-situ snow particle measurements, whether ground-based or airborne, are highly valuable but are typi-

30 cally sparse and often insufficient to capture the complex spatio-temporal evolution of the particles. Besides, certain quantities like particle mass are particularly difficult to measure and usually available only for small sets of particles, although recent technical and methodological developments open up the possibility for automatized estimations (Leinonen et al., 2021; Rees et al., 2021).

Alternatively, remote sensing instruments, such as meteorological radars, provide measurements related to the scattering

35 of an electromagnetic signal by an ensemble of hydrometeors, over a vertical column of the atmosphere for profiling radars or full 3D regions for scanning ones. However, such measurements are indirect and it is needless to say that the retrieval of snow microphysics from radar variables is not explicit. Due to the large variability of snow crystal geometrical and scattering properties, too strong simplifications of the radiative calculations may yield erroneous results (Leinonen et al., 2012). Recent research efforts in this direction have brought about significant improvements in scattering models (e.g. Kuo et al., 2016; Lu

40 et al., 2016; Hogan et al., 2017; Ori et al., 2021). In spite of this progress, the estimation of microphysical properties from radar measurements often remains an ill-posed problem, and is further hindered by measurement uncertainty, for example related to instrument miscalibration or attenuation along the radar path.

Radar retrievals and analyses of snowfall microphysics have been successfully conducted using two distinct approaches. On the one hand, the use of multi-frequency measurements has become quite popular: this approach relies on the fact that large

45 snow particles transition to non-Rayleigh scattering regimes at millimeter wavelengths, while they remain Rayleigh scatterers at larger wavelengths. Combining measurements from a shorter and a longer wavelength radar (e.g. W- and X-band), the dual-frequency ratio of radar equivalent reflectivity ($DFR = Ze_X - Ze_W$, in dB) can thus be used to identify populations of snow particles with a larger size or density (e.g., Matrosov et al., 1992; Matrosov, 1998; Szyrmer and Zawadzki, 2014; Liao et al., 2016), thus indicating regions of enhanced snowfall growth. With three well-chosen radar frequencies, studies were

50 able to identify distinct signatures for riming and aggregation mechanisms, and even retrieve estimates of fractal dimension during some parts of snowfall events (e.g. Kneifel et al., 2011; Kulie et al., 2014; Leinonen et al., 2018a), which were later comforted through comparison with in-situ airborne data (Nguyen et al., 2022). Similar retrievals, focusing on snow density, were achieved using only two frequencies, but leveraging information contained in the mean Doppler velocity in addition to radar reflectivity (Mason et al., 2018). Bringing this a step further, Mroz et al. (2021) were recently able to retrieve from triple-

55 frequency radar reflectivity and Doppler velocity accurate estimates of ice water content, snow particle characteristic size as well as an estimate of riming degree. In the case of scanning radars, additional polarimetric information can be included which opens up new possibilities, e.g. for the geometrical description of ice-phase hydrometeors (Tetoni et al., 2021). Most of the

cited studies however rely on certain hypotheses, for example on the mass-size relation, with assumptions ranging from the use of a strict parameterization to more flexible yet still constraining models like the "filling-in" hypothesis (Mroz et al., 2021).

60   On the other hand, more qualitative studies have been conducted relying not solely on radar moments (e.g. reflectivity Z_e or mean Doppler velocity MDV) but rather on the full Doppler spectrum. By observing wider, more skewed, or even multi-modal spectra, signatures of specific microphysical processes can be identified such as riming or aggregation (e.g. Shupe et al., 2004; Kalesse et al., 2016). Indeed, the full Doppler spectrum encloses more information on microphysical properties and the particle size distribution (PSD) than scalar moments like Z_e or MDV.

65   Combining multi-frequency and Doppler spectral techniques appears like a promising way to go, possibly allowing to reduce the number of required assumptions for a microphysical retrieval, as investigated by Kneifel et al. (2016) and Barrett et al. (2019). The scattering regime transition in high frequencies is in principle visible in dual-frequency Doppler spectra with the following signature: slow-falling particles are typically Rayleigh scatterers and contribute to similar reflectivity at both wavelengths; while larger, fast-falling particles are no longer Rayleigh scatterers for the higher frequency with thus smaller spectral reflectivity than for the lower frequency. This means that the Doppler spectra at both frequencies should "match" on the low-velocity side, and diverge for large velocities. However, using this principle to perform a direct inversion like Barrett et al. (2019) is only rarely possible. Difficulties related to imperfect measurements are substantial: not only should the different radars be well cross-calibrated in reflectivity, they should also be well aligned vertically to avoid contamination by horizontal wind; beam mismatching caused by the difference in beam widths can also hinder the retrieval, especially when turbulent broadening is observed, or when the particle populations in the sampled volumes are too heterogeneous. In such cases, a direct computation of the dual-frequency spectral ratio is meaningless.

In this work, we propose a novel approach to retrieve snowfall microphysics from dual-frequency Doppler spectra, while partly relaxing these constraints on turbulence or beam matching, as well as reducing the number of prior assumptions on snowfall microphysical properties. Whereas many retrievals in atmospheric sciences rely on classical Bayesian frameworks (e.g. Rodgers, Clive D., 2000), we opt here for an alternative machine-learning based method: some cutting-edge developments achieved in the past decade have outlined the strong potential of such statistical methods in atmospheric science (Bauer et al., 2021; Chantry et al., 2021) and weather radar applications (Geng et al., 2021), especially to tackle retrieval problems (Vogl et al., 2022; Chase et al., 2021).

Exploiting recent advances in deep learning research, we introduce a novel physics-constrained inversion framework, which is partly inspired from auto-encoder models. The auto-encoder is a neural network architecture originally designed for dimension reduction purposes, and sometimes referred to as a non-linear principal component analysis variant (Kramer, 1991; Hinton and Salakhutdinov, 2006): an *encoder* neural network maps a high-dimensional signal to a low-dimensional *latent* or *feature* space, while the *decoder* neural network learns to recover the original signal from this latent space. In our case, dual-frequency Doppler spectrograms constitute the high-dimensional signal, while the dimensions of the latent space are constrained to represent the snowfall properties which we seek to retrieve. In a first step, the decoder is trained to emulate a radiative transfer model, i.e. to reconstruct dual-frequency Doppler spectrograms given (latent) snowfall descriptors. The encoder is trained in a second step: it consists of an advanced deep learning architecture, that ingests the radar data (dual-frequency Doppler spectrograms),

and is optimized to retrieve the latent snowfall properties which, when passed through the decoder, minimize the reconstruction error with respect to the input data. An important peculiarity of the encoder's architecture is its ability to leverage the spatial

95   consistency of the radar variables, which reduces the ill-posedness of the inversion problem.

The proposed framework is implemented on data from the ICE GENESIS campaign that took place in the Swiss Jura in January 2021. The set-up included in particular X- and W-band Doppler spectral profilers at the ground, complemented with overpasses of a scientific aircraft equipped with microphysical probes. This offers the possibility to validate the retrieval against in-situ measurements.

100   A general overview of the retrieval framework and its theoretical foundation is presented in Sect. 2. Section 3 is dedicated to the presentation of the synthetic and real data sets used to train and evaluate the inversion model. In Sect. 4, we detail the technical implementation of the framework. Results are then presented in Sect. 5, with a particular focus on the comparison of the retrieval outputs to in-situ aircraft measurements. The discussion of the results is taken a step further in Sect. 6, with a focus on the method's sensitivity and its current limitations.

105   ## 2   Theoretical framework

This section introduces the theoretical components required to understand the proposed retrieval framework, and provides an overview of its general structure.

2.1   Doppler spectra: forward model

Radar Doppler spectra are computed through the Fourier transform of the radar return signal (Doviak and Zrnic, 1993); they

110   feature the reflectivity-weighted distribution of the targets' Doppler velocity in a given radar volume. From here on, *"Doppler spectrum (pl: spectra)"* refers to the measurement at a given time and range gate, and the vertical stack of spectra at a certain time is denoted as the *"Doppler spectrogram"*. Note that this name convention is used here for clarity, but it may not be universal.

The shape of a Doppler spectrum in snowfall results from a combination of several factors, which are mainly the snowfall mi-

115   crophysical properties —e.g. PSD, bulk density, geometry, etc.—, atmospheric conditions —turbulence, horizontal and vertical wind— as well as the radar parameters —sensitivity, beam width, frequency. Understanding how those parameters (microphysical, environmental, instrumental) translate into a measured Doppler spectrum is delicate: it involves complex radiative transfer models to compute the radar backscatter of snow particles, and it also requires an understanding of snowfall aerodynamic properties e.g. for the parameterization of the particle velocity-size relations.

120   Efforts have been devoted to the construction of increasingly accurate forward models, for instance through computationally costly discrete dipole approximation (DDA) calculations (e.g. Draine and Flatau, 1994; Liu, 2004; Kuo et al., 2016; Lu et al., 2016), and through simulations based on the self-similar Rayleigh-Gans approximation (SSRGA), which have been tuned to represent accurately the scattering of various particle types (Hogan and Westbrook, 2014; Hogan et al., 2017; Ori et al., 2021). In this work, we use as a forward model the radiative transfer code PAMTRA (Mech et al., 2020), which is particularly suited

125   to simulate full Doppler spectra and offers an implementation of several scattering models. Details on how PAMTRA is used
and parameterized in this study are presented in Section 3.1.1

2.2   Approach to the inverse problem

Assuming the forward model, noted $\mathbf{f}$, is now known —which, given a set of properties x, outputs realistic Doppler spectra
y— the aim of the retrieval is to solve the following *inverse problem*: from real observed spectra y_r, estimate the underlying

130   microphysical properties x_r (see e.g. Maahn et al., 2020, for a discussion on inverse problems). Here the subscript r denotes
real values as opposed to synthetic or modeled quantities.

In a mathematical language, this means estimating $\mathbf{g} = \mathbf{f}^{-1}$. This is in general not possible, because $\mathbf{f}$ is usually not an
invertible mapping. Workarounds can be developed in certain cases, for example through look-up tables (e.g Leinonen et al.,

135   2018b). Alternatively, one can seek x_r as the minimizing argument of a cost function (e.g. $||y_r - \mathbf{f}(x)||^2$), which can also
include a regularization term (e.g. Mason et al., 2018); this minimization problem can then be solved iteratively with for in-
stance a gradient descent algorithm. From a Bayesian perspective and under additional assumptions (e.g. Gaussian probability
distributions), this corresponds to the popular *Optimal Estimation* (OE) framework (Rodgers, Clive D., 2000; Maahn et al.,
2020), which is widely used across atmospheric science to solve moderately linear inverse problems. Although this alleviates

140   some requirements on $\mathbf{f}$, it can only be implemented if $\mathbf{f}$ is differentiable, and if the computation of its gradient is tractable,
either analytically or numerically.

This "classical Bayesian" approach faces some limitations, which include but are not limited to, the need to linearize the for-
ward operator in order to compute its Jacobian, or to assume prior values for x. This, combined with the complexity of our
forward model (i.e. PAMTRA), has led us to explore a different direction.

145

Machine learning techniques offer the possibility to tackle inverse problems in a different way, i.e. with a statistical rather
than an analytical approach. Note that, as pointed out by Geer (2021), the overarching framework in both cases ultimately
remains that of Bayesian probabilities, viewed through different prisms. The typical machine learning route to solve an inverse
problem (e.g. Chase et al., 2021) has the following structure: the available forward model is first used to create a large synthetic

150   dataset $\{(x_s^k, y_s^k = \mathbf{f}(x_s^k)), k = 1..N\}$; the s subscript denoting synthetic values and N the size of the dataset. Then, a machine
learning model is trained on this dataset to learn a statistical relation between y and x, i.e. an approximation of the inverse
mapping $\tilde{\mathbf{g}}$. Ultimately, this produces a gate-to-gate inversion of the problem which can be implemented on real data.

One major limitation of this "direct" method is when the problem itself is ill-posed, i.e. when several values of x may yield
similar outputs y: in such cases, the retrieval may yield arbitrary outputs.

155   The proposed approach, illustrated in Fig. 1, can mitigate this issue.

– In a first step, a neural network is trained on a synthetic dataset of $(x_s, y_s = \mathbf{f}(x_s))$. Instead of learning an inverse
mapping, it simply learns to emulate the forward model: taking microphysical and atmospheric (i.e. *latent*) variables (x_s

as input, it outputs Doppler spectra (y_s. This model, which we hereafter refer to as the *"decoder"* and denote $\tilde{\mathbf{f}}$, is thus a differentiable emulator of PAMTRA. When applied not to a single set of microphysical descriptors, but to a stack (e.g. of multiple range gates) at once, it is denoted with upper-case $\tilde{\mathbf{F}}$. This synthetic dataset should include a wide range of realistic parameters, to not induce bias in the further steps.

– In a second step, we shift our attention to a real (i.e. not synthetic) dataset of full dual-frequency Doppler spectrograms Y_r; the aim is to retrieve the underlying profiles of latent variables X_r. The capital letter denotes that e.g. Y_r is a vertical stack of y_r. A second neural network, the encoder $\tilde{\mathbf{G}}$, is trained on this real data set: it takes as input the spectrograms Y_r, and its output $X = \tilde{\mathbf{G}}(Y_r)$ has the same dimension as the number of latent features, times the number of range gates. X is passed on to the decoder $\tilde{\mathbf{F}}$, which outputs a reconstructed spectrogram Y. Training is performed by optimizing $\tilde{\mathbf{G}}$ in order to minimize the reconstruction error $||Y - Y_r||^2$.

At the end of the training i.e. when the pipeline has converged, $\tilde{\mathbf{F}} \circ \tilde{\mathbf{G}}(Y_r) \approx Y_r$ and the output of the encoder $\hat{X}_r = \tilde{\mathbf{G}}(Y_r)$ should be close to the true profile of microphysical descriptors X_r.

The architecture of the decoder and encoder will be detailed further on (Sect. 4), but one key property should already be underlined. The encoder network contains convolution kernels: thanks to this feature, it can capture the spatial structure of the Doppler spectrograms, e.g. their vertical continuity, the presence of shear layers, etc. The network is thus able to preserve the spatial consistency of the measurements, and to output profiles which are themselves spatially consistent, thereby reducing the ill-posedness of the problem. We believe that this implicit use of the measurements' spatial features in the retrieval is a key contribution of this work. To support this, a brief discussion of alternative methods is proposed further on (Sect. 6.4.

To conclude this overview of the framework, we highlight that while it was presented for the specific case of Doppler spectrograms and snowfall microphysics, its structure is generic and could potentially be applied to other retrieval problems with similar properties: a complex forward model that is not directly invertible, with slightly ill-posed features that hamper pointwise retrievals.

3 Data

3.1 Synthetic data set

As mentioned above (Sect. 2 2), the first step of the framework consists in training a neural network on a synthetic dataset containing sets of microphysical and atmospheric variables and the corresponding spectra. The focus of this section is the generation of this training set.

3.1.1 Forward model assumptions

To generate this synthetic dataset, PAMTRA is run by prescribing snowfall microphysics through several parameters. These parameters are the snowfall properties that the algorithm will then learn to retrieve.
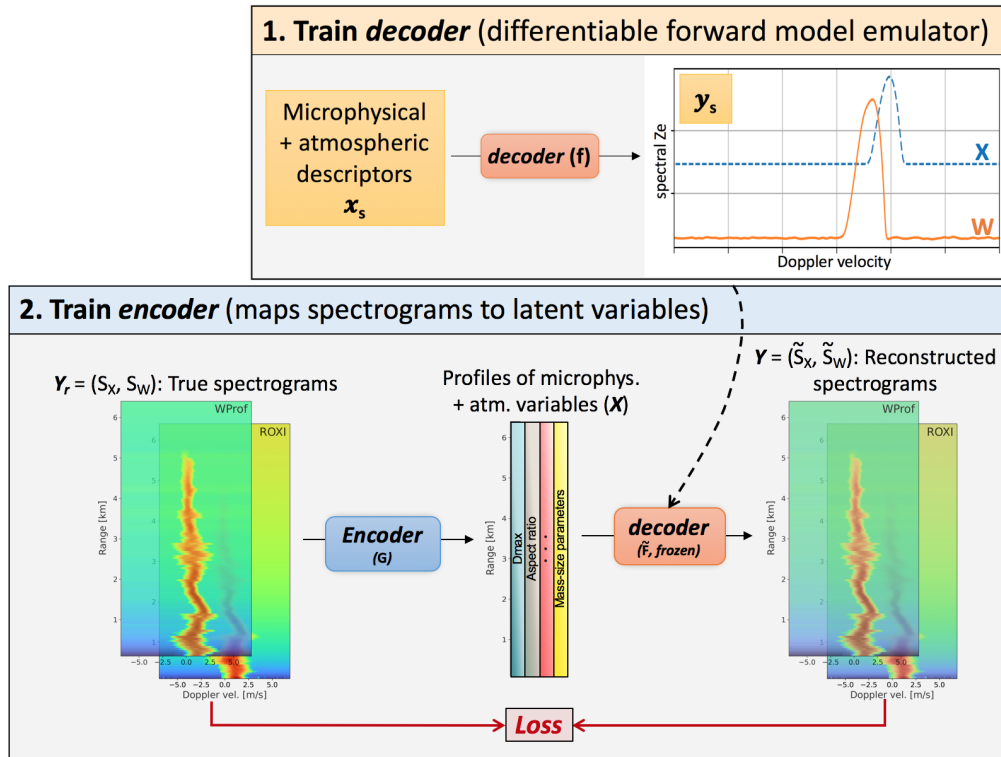
Figure 1. Schematic illustration of the method.

The definitions of the microphysical parameters are summarized in Table 1. The PSD is assumed to be a negative exponential

190 $(N(D) = N_0 \exp(-D/D_0)$, e.g. Straka, 2009), whose size parameter D_0 is prescribed; here and further, the size of a particle is defined as its maximum dimension: $D = D_{max}$. For an exponential PSD, D_0 is equal to the mean diameter; the effective diameter, often relevant for radiative transfer models, is $D_{eff} = 3D_0$. Mass-size and area-size relations are considered to be power laws, whose prefactors and exponents are prescribed ($m = a_m D_m^b$, $A = \alpha_a D^{\beta_a}$). Aspect ratio A_r is then specified, as well as the ice water content (IWC). Note that the particle number concentration is implicitly prescribed through the definition

195 of IWC, D_0 and a_m, b_m. The noise level is additionally specified, since it is required to simulate Doppler spectra; in practice, it only depends on the range and on the radar's properties, and is not related to other microphysical or atmospheric quantities. The velocity-size relation is the one proposed by Heymsfield and Westbrook (2010).

Individual spectra are simulated through PAMTRA for an altitude of 1000 masl, using a standard (PAMTRA default) atmospheric profile with a temperature randomly chosen in [-20 °C, 1°C]. Scattering calculations are performed using the SSRGA,

200 with coefficients from Ori et al. (2021); more detail on this is provided in Appendix B2. These assumptions on scattering properties are not flawless, and constitute a bottleneck in our method, as in virtually any attempt at radar-based retrievals. In particular, the current implementation of PAMTRA (28/03/2022) allows for the parameterization of only two coefficients ($\kappa_{SSRG}, \beta_{SSRG}$), while current literature suggests that more coefficients should be used (γ_{SSRG} and ζ_{SSRG}, cf. Hogan et al.

(2017) for detail on the coefficients); furthermore, the variability of the scattering parameters shown e.g. in Ori et al. (2021) or
205 Leinonen et al. (2018a), Fig. 5, was overlooked when the parameters were sampled (cf B2). These assumptions can naturally be
questioned. We however believe that it was reasonable to use the simplest possible parameterization for the initial development
of the method, leaving to future studies the possible improvements of the forward model.

3.1.2   Forward model inputs

When generating this training set, a trade-off has to be defined: if the dataset is too narrow, i.e. if it doesn't cover a large enough
210 range of values and combinations for the microphysical descriptors, this will cause a bias in the retrieval; conversely, if the range
of values is much too large, this will hinder the training process, for it will include non-realistic values. It was therefore chosen
to parameterize PAMTRA by sampling the microphysical properties using a large observational dataset collected using the
Multi-Angle Snowflake Camera during 10 field deployments. This data was organized into a database in Grazioli et al. (2022).
We follow the method presented in this study (Grazioli et al., 2022, Sect. "Technical Validation") to derive from the database
215 the microphysical parameters required in the forward model (Table 1). Four categories of particles are used: aggregates, planar
crystals, columnar crystals and graupel. For each type of particle and for each parameter, a distribution is fitted to the empirical
histogram calculated from the database. We refer to Appendix B1 for more detail. When generating the training set, parameters
are then randomly sampled from those distributions. It is worth highlighting that all parameters are sampled independently,
with the exception of a_m and α_a. Indeed, as pointed out in Grazioli et al. (2022), a strong correlation exists between a and b,
220 and between α_a and β_a; thus, empirical fits are used from which a_m (resp. α_a) is sampled for a given b_m (resp. β_a), with the
addition of randomness using the mean squared error of the fit. The definition of aspect ratio A_r is slightly different between
the MASC dataset and the SSRGA parameterization: in the former, it is equal to the ratio of minor axis length to major axis
length (Garrett et al., 2015), while in the latter it is equal to the particle's dimension along the direction of radar beam (here,
vertical) divided by maximum dimension (Ori et al., 2021). After some empirical exploration, it was decided to use nonetheless
225 the histograms from MASCDB, given that the distributions were quite broad, meaning this difference in definition should not
bias the retrieval. Ice water content is the only parameter for which the MASCDB does not provide estimates; therefore, it was
empirically decided based on literature (Noh et al., 2013) and preliminary analyses of aircraft in-situ measurements during the
ICE GENESIS campaign (cf Sect. 3.2) to sample it from a negative exponential distribution with a mean of $0.5 \, \mathrm{g \, m^{-3}}$.

3.1.3   Generation of the training set

230 Each item of the dataset is generated through the following procedure.

1. A particle type is randomly sampled among the four aforementioned types. Given the large variety that exists within the
   aggregate category, it is given more weight in the sampling procedure (aggregates: 40% - planar crystals: 20% - graupel:
   20% - columnar crystals: 20 %).

2. Microphysical descriptors are randomly sampled using the MASC-based distributions.

Table 1. Microphysical, atmospheric and radar parameters

Name	Description
IWC	Ice water content
D_0	Mean diameter (assuming exponential PSD)
b_m	Exponent of the mass-size power law
a_m	Pre-factor of the mass-size power law
β_a	Exponent of the mass-size power law
α_a	Pre-factor of the mass-size power law
A_r	Aspect ratio (cf. Sect. 3.1.3)
$TurbX$	Broadening X-band
$TurbW$	Broadening W-band
$WindX$	Radial wind X-band
$TurbW$	Radial wind W-band
$LnoiseX$	Noise level at X-band
$LnoiseW$	Noise level at W-band

3. PAMTRA is run on these descriptors, under the previously stated assumptions. The corresponding Doppler spectra are computed for 9.48 and 94 GHz (frequency of the radars used in this study, cf. Sect. 3.2), with 512 bins and a Nyquist velocity of $6.92\,\mathrm{m\,s^{-1}}$.

4. Then, turbulent broadening and spectrum shift due to radial wind are added, with randomly sampled values. While these could be computed directly in PAMTRA, it was more computationally efficient to implement them in post-processing in a vectorized way. The radial wind parameter includes the wind shift that could be caused e.g. by vertical wind, beam misalignment, etc. The broadening parameter is the size of the Gaussian broadening kernel and includes the effect turbulent eddies but in reality also accounts for all other possible broadening causes (e.g. horizontal wind, beam width). It is computed by sampling a value of atmospheric turbulence, represented by the eddy dissipation rate (sampled as a negative exponential with $10^{-3}\,\mathrm{m^2s^{-3}}$ mean, consistent with some literature standards, e.g. Sharman et al. (2014)). The resulting broadening is derived following equations of Shupe et al. (2012) using for the radar's beam width those of the W- and X-band radars used in the study, described in Section 3.2.

5. Finally, spectra are reduced to 256 points through bin averaging, for computational reasons.

Ultimately, each item of the synthetic dataset contains an input vector with 13 dimensions (see table 1) and the corresponding simulated Doppler spectra (X- and W-band) with each 256 bins. We underline that the synthetic data set contains information only at the scale of the radar sampling volume at a given range gate, i.e. not a full spectrogram.

3.2 X- and W-band Doppler spectrograms

In this section, we present the experimental dataset used for the implementation of the second part of the pipeline. Measurements were collected during the 2021 ICE GENESIS campaign, a joint ground-based and airborne field experiment that was conducted in the Swiss Jura in January 2021, and is fully described in Billault-Roux et al. 2022 (under review). Data from X-255 and W-band vertically-pointing Doppler radars are used, which were located at les Eplatures airport. The X-band radar, further on referred to as ROXI (Viltard et al., 2019), is a high-sensitivity 9.48 GHz Doppler spectral profiler with 1.8° 3-dB beam width. It was deployed next to a dual-polarization W-band 94 GHz Doppler spectral profiler (WProf, Küchler et al., 2017) with 0.53° beam width. The properties and settings of both radars are summarized in table 2. Radar cross-calibration, as well as attenuation issues, are discussed in Appendix A.

Table 2. Parameters of ROXI and WProf radars during the ICE GENESIS deployment. WProf range and time resolution (as well as Nyquist velocity) are defined in three chirps. The lower chirp ranges from 100 m to 900 m, the second one from 900 m to 3900 m and the higher one from 3900 m to 9000 m

Radar properties	ROXI	WProf		
		chirp 0	*chirp 1*	*chirp 2*
Frequency (GHz)	9.48	—	94	—
Beamwidth (°)	1.8	—	0.53	—
Time resolution (s)	5	—	5	—
Range resolution (m)	50	7.5	16	32
Velocity resolution (m/s)	0.1	0.02	0.014	0.013
Nyquist velocity (m/s)	11	10.8	6.92	3.3

260 Then, the spectrograms of both radars are remapped to a common grid, by averaging in time (with a resolution of 20 s), interpolating in range (resolution of 50 m), and average-binning the velocity to the same bins as the synthetic dataset, i.e. with 256 bins and a velocity cutoff $v_{Nyq} = 6.92 \ \mathrm{ms}^{-1}$.

3.3 Data for model evaluation

3.3.1 Polarimetric radar

265 MXPol (Schneebeli et al., 2013) is a polarimetric X-band scanning radar that was deployed 4.8 km away from the main site and that performed routine range-height indicator (RHI) scans in direction of the X- and W-band profilers during precipitation. Hydrometeor classification with demixing (Besic et al., 2016, 2018) was performed on this data to estimate from the polarimetric variables the proportions of hydrometeor types in the sampled volume. From the RHIs, remapped to a Cartesian grid, profiles are extracted over the main site with a horizontal $\delta x = \pm 500$ m, using only elevation angles below 45 degrees. The

time series of hydrometeor classification extracted in this manner will be used qualitatively as an independent verification tool to assess the performance of our microphysical retrieval.

3.3.2 Aircraft in-situ measurements

In addition to the ground-based measurements, the ICE GENESIS campaign included scientific aircraft overpasses with remote-sensing and in-situ instruments. The airborne in-situ data is particularly valuable for the quantitative evaluation of the micro-physical retrieval, presented in Sect. 5. Airborne measurements used in this work were collected during three flights of the Safire ATR-42 (Jan. 22, Jan. 23, Jan. 27) as the aircraft was performing overpasses over the ground site, with data from several probes.

First, the Counterflow Virtual Impactor (CVI, Anderson et al., 1994; Schwarzenboeck et al., 2000) provides a measurement of total water content (TWC), and the Cloud Droplet Probe (CDP) of liquid water content (LWC); from those measurements, an estimate of IWC can be obtained as $IWC = TWC - LWC$. Two imaging probes are also used, the 2D-Stero probe (2D-S) and the Precipitation Imaging Probe (PIP) sampling respectively 10 μm to 1.28 mm and 100 μm to 6.4 mm (Baumgardner et al., 2017; McFarquhar et al., 2017, for a complete reference). From the images of these probes, the method of Leroy et al. (2016) is used to is compute the PSD and derive the following microphysical descriptors: aspect ratio, mass-size power law coefficients, area-size power law coefficients. For the mass-size parameters, in addition to the CVI closure method proposed by Leroy et al. (2016), another method is used for particle-by-particle mass reconstruction based on Lawson and Baker (2006); the methods are respectively denoted with the CVI and BL subscripts.

3.3.3 Airborne radar retrieval

The aircraft was also equipped with an upward-looking W-band radar (RASTA, Plana-Fattori et al., 2010) from which values of IWC were derived (Delanoë et al., 2007), noted with the $RASTA$ subscript. In order to compare IWC_{RASTA} to the in-situ measurements, the closest valid radar gates are used, i.e. at a vertical distance of 150 to 250 meters above the aircraft. For a fair comparison between airborne RASTA retrievals and our inversion model, only time steps when the aircraft overpasses the ground site are used, i.e. when the aircraft is within a 1 km horizontal distance to the ground site (distance chosen to allow a sufficient number of points for the comparison).

4 Deep learning inversion framework

This section addresses the detail of the implementation of the two-step framework outlined in Sect. 2. For designing and training both the decoder and the encoder part of the model, the pytorch library is used (Paszke et al., 2019).

4.1 The "decoder": a differentiable emulator of PAMTRA

The first part of the framework consists in developing a differentiable emulator of PAMTRA, by designing a deep learning model and training it on the synthetic dataset (Sect. 3.1). If viewed in perspective with the technique of auto-encoders, this

300   consists in learning the "decoder", which maps the latent space —containing the physical variables— to the high-dimensional measurement space —the spectra. It was chosen to train separately the X-band and W-band decoders rather than use a single algorithm emulating both simultaneously; indeed, the two frequencies may have slightly different smoothness or amplitude features, which justifies the use of distinct architectures. Each decoder takes as input a vector of dimension 10 containing IWC, D_0, b_m, a_m, β_a, α_a, A_r, $Turb_F$, $Wind_F$ and $Lnoise_F$, where F is either X or W. They output Doppler spectra with

305   256 points.

4.1.1   Decoder architecture

The model, whose architecture is illustrated in Fig. 2 a), is designed as a neural network (NN) with a first part composed of fully-connected layers, and a second part composed of convolutional layers. More precisely, since the aim of this decoder is to increase dimensionality (from 10 to 256), we use one-dimensional "transposed convolutions" which are well suited for this

310   purpose (Zeiler et al., 2010). Since using only transposed convolutions can create artifacts, another type of layer is also used to ensure a smooth output, i.e. a linear upsampling layer followed by a standard convolution. In order to improve the training of the model, *residual blocks* (He et al., 2015) are used: these blocks contain skip connections and batch normalization steps (Ioffe and Szegedy, 2015), and are quite popular in deep learning applications. In a nutshell, these techniques help mitigate issues caused by the depth of the model: they do not per se improve the expressiveness of the neural network, but they strongly

315   facilitate the training process.

4.1.2   Decoder training

The NN is trained on the synthetic dataset described in Sect. 3.1, which is split into training, validation and testing sets (80% - 10% - 10%). The input is normalized using the statistics of the training dataset, i.e. with the mean and standard deviation of each variable. For certain variables, the natural logarithm is used instead of the original value, in order to have more

320   homogeneously spread distributions: this is the case for IWC, a_m and α_a. The network is trained using the Adam optimizer (Kingma and Ba, 2015), with mean square error (MSE) as a loss metric, and with Xavier normal initialization of weights and biases (Glorot and Bengio, 2010); in addition, the learning rate is periodically decreased with a scheduler. The network and training hyperparameters are summarized in Table 3. It was observed that the spectra output by the NN could have a tendency to slightly underestimate the peak values, because of a "flattening" effect common in such methods that use MSE as a loss

325   metric. Hence, we add to the main loss a secondary loss calculated as the mean square error computed only on the part of the spectrum close to its peak (above 50 % of its amplitude).

4.2   The "encoder": retrieving a profile of latent variables

In the second part of the framework, a second deep neural network, the *encoder*, is used to learn the inverse mapping. Taking as input dual-frequency spectrograms, i.e. an array of shape $N_{rg} \times 256 \times 2$, with N_{rg} the number of range gates, it outputs

330   vectors in the latent space, of shape $N_{rg} \times 13$, 13 being the total number of latent dimensions retrieved, as detailed in Table 1.
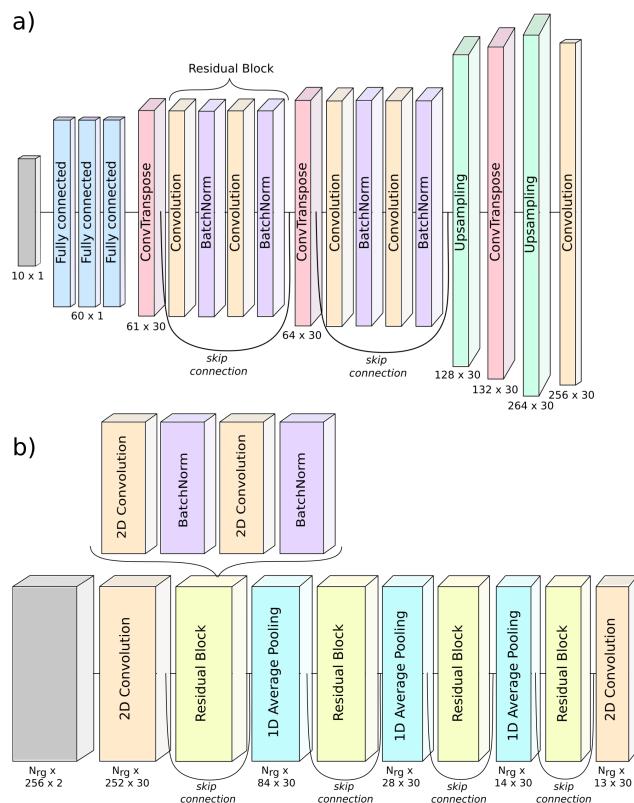
Figure 2. Architecture of the decoder (a) and encoder (b) neural networks. For the decoder, the architecture for W-band is shown; that of X-band is extremely similar with only slightly different kernel widths. Skip connections indicate that the output of a given layer is kept and added further on to the output of a residual block. Color coding indicates the type of each layer. The size of each layer is indicated; when not stated, the layer is the same size as the previous one. Note that for display reasons the velocity dimension is along the vertical in a) and along the horizontal direction in b).

In this section, we refer to the first dimension as the "range" dimension, to the second one as the "feature" dimension and the last one as the "channel" dimension.

4.2.1 Encoder architecture

The neural network designed for this part uses 2-dimensional convolutions, which allow to reduce the feature dimension from

335   256 to 13. In order to keep the range dimension constant (equal to N_{rg}), padding is used, which means artificially increasing the size of the array —by replicating on each side the items resp. in first and last position— before performing the convolution. Similar to the decoder architecture, residual blocks with skip connections and batch normalization steps are used. Additionally, average pooling along the feature dimension is performed after each residual block. The size of the convolution kernels along

the range dimension reflects the scale at which we can expect meaningful spatial consistency —both in the latent space and in
340   the measured spectrograms. After some exploration, it was fixed to 3, i.e 150 m of range. The full architecture of the encoder
is displayed on Fig. 2 b).

4.2.2   Encoder training

The encoder is trained using the radar data presented in Sect. 3.2. The spectrograms are normalized using the statistics used
for normalization in the decoder part (means and standard deviations from the synthetic spectra). Rather than using the entire
345   Doppler spectrograms ($N_{rg} = 100$) as one training item, chunks of the spectrograms are used with $N_{rg} = 25$, and are sampled
in the following way: the first chunk corresponds to $i_{rg} = 0..24$, the second chunk to $i_{rg} = 5..29$, the third to $i_{rg} = 10..34$, etc.;
i_{rg} being the range gate index. The dataset is then randomly shuffled at each epoch during training. Rearranging the dataset
this way makes training both more tractable —thanks to the smaller size of each item— and more robust —due to the data
augmentation, which helps avoid local minima during the training process. During training, the encoder output is passed as
350   input to the X- and W-band decoders, which then output reconstructed spectrograms. The reconstruction loss is the MSE be-
tween the reconstructed ($\tilde{S}_W, \tilde{S}_X$) and the original spectrograms (S_W, S_X): $Loss = (\tilde{S}_X - S_X)^2 + (\tilde{S}_W - S_W)^2$. The encoder
parameters are then updated at each step to minimize this loss, here with the Adam optimizer. Training parameters are reviewed
in Table 3. It is important to note that the decoders' parameters are frozen at this step: only the encoder is being learned; this
differs from classical autoencoder models, for which the decoder and encoder are trained simultaneously.

355

Estimates of the latent features are then obtained from the encoder's output, after inverse normalization and exponential
transform for those variables whose logarithm was used as input to the decoder (see Section 4.1.2). To prevent occasional
convergence toward unrealistic values in the latent space (e.g. $D_0 < 0$), an additional constraint was incorporated to the loss
term to penalize latent values outside of a manually defined range: for a given feature x with realistic bounds x_{min} and x_{max},
360   this secondary loss reads $L_{sec}(x) = \mathbb{1}_{(x<x_{min})} \times (x - x_{min})^2 + \mathbb{1}_{(x>x_{max})} \times (x - x_{max})^2$.

4.3   Note on uncertainty quantification

In order to get an estimation of the retrieval's uncertainty, an ensemble approach is used: several runs are performed for both
the decoders and the encoder, each trained independently with random weight and bias initialization. In the end, a total number
365   of 50 runs is used to compute mean values and standard deviations for each retrieved variable. This both ensures a greater
robustness of the retrieved values, which are less likely to reflect local minima, and provides an uncertainty estimate for the
retrieval. On the downside, this implies a lengthier process since training is a computationally demanding task that typically
takes a few hours on a standard GPU.

Table 3. Hyperparameters of the encoder and decoder neural networks.

Hyperparameter	Decoder	Encoder
Activation	ReLU	ReLU
Range of kernel sizes	$1 \times (2-7)$	$3 \times (3-7)$
Number of channels	30	30
Number of linear layers	3	0
Number of neurons in linear layers	60	-
Total number of parameters	35000	150000
Padding mode	*replicate*	*replicate*
Loss	MSE	MSE
Optimizer	Adam	Adam
Batch size	250	15
Number of epochs	3	200
Learning rate (initial)	1e-3	1e-3
Optimizer epsilon	1e-8	1e-6
Scheduler step / rate	0.6 / 0.2	90 / 0.5
Parameter initialization	Xavier	Xavier

5   Results

5.1   Training convergence and accuracy

This section is dedicated to the evaluation of the pipeline and the verification of its convergence, which is a necessary step before examining the retrieved latent variables themselves.

5.1.1   Decoder

The training of the decoder networks was successful, with the loss function decreasing with the number of epochs until it plateaus. It was verified that increasing the training set size did not result in a change of this plateau value, meaning that the training dataset was large enough for the chosen neural network's complexity. Examples of model outputs are shown in Fig. 3.

Since the loss function, i.e. MSE on normalized spectra, is not easily interpretable to assess the model's performance, the following overlap metric was defined: $O(S, \tilde{S}) = 0.5 \frac{\int min(S^*, \tilde{S}^*)}{\int S^*} + 0.5 \frac{\int min(S^*, \tilde{S}^*)}{\int \tilde{S}^*}$ where $S^* = S - min(S)$ and $\tilde{S}* = \tilde{S} - min(S)$; $O(S, \tilde{S})$ is equal to 1 (or 100 %) when the spectra are perfectly identical, and to 0 if they are disjoint; in the formula, S^* and $\tilde{S}^*$ are introduced to bring the noise level of the spectra to 0 and thus ensure that the overlap is computed on the signal only.

On the testing set, the overlap for X-band (resp. W-band) is of 90.7% (resp. 94.8%), as an average over 5 different runs with random initialization. This reflects a good, although not perfect, performance of the algorithm. Looking at a few examples of individual spectra, it comes across that the model has a slight tendency to underestimate the peak of the spectrum (see for

385 example the X-band spectrum chosen in Fig. 3), despite the secondary loss that was used; however, the rather good overall agreement between target and output spectra suggests this is not a prohibitive issue. Additionally, we observe that performance is slightly worse for X than for W-band spectra, in spite of some efforts to adjust the neural network architectures independently to improve each model's accuracy. This could be related to the fact that W-band spectra have a lower noise level, meaning the actual signal —the peak— occupies a larger part of the spectrum than for X-band; this could in turn facilitate learning.
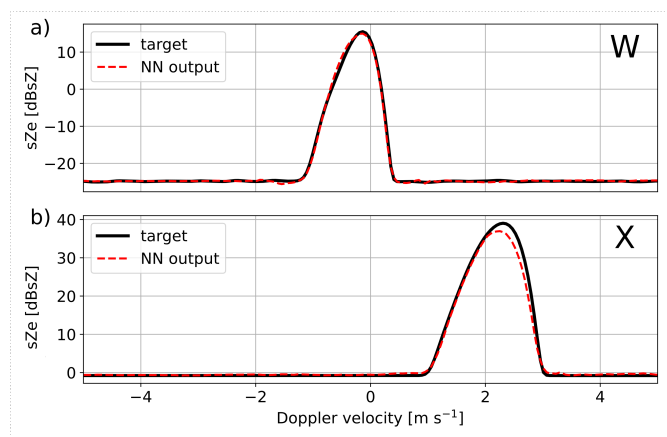
Figure 3. Examples of testing set results for W- and X-band retrievals, with model output (dashed red) and target spectrum (black line). The examples were chosen to reflect some of the typical behaviors and possible artifacts that were observed. Units of spectral reflectivity are used, i.e. $1\,\mathrm{dBsZ} = 10\log_{10}(1\,\mathrm{mm}^6\mathrm{m}^{-3}(\mathrm{ms}^{-1})^{-1})$. Doppler velocity is positive for downward motion.

5.1.2 Encoder

390

Training of the encoder is also successful: the full pipeline is able to reconstruct original spectrograms in a satisfactory manner, as is visible in Figure 4. Only W-band spectrograms are shown but results are visually very similar at X-band. The overlap metrics are slightly below the ones of the decoder alone (respectively 86% and 91% for X- and W-band spectrograms); this slight decrease can be expected for several reasons: first, the real spectrograms include high-shear regions with significant

395 turbulent broadening (which can be visually identified as regions with suddenly much wider spectra, along with variable velocity, e.g in Fig. 4 below 500 m), which the model cannot be expected to resolve perfectly. Then, some time steps include bimodal spectra (e.g. 2.5 km - 3.5 km) which the model in its current state is unable to replicate. When looking only at spectrograms with moderate apparent turbulence (e.g. 0.5 km - 1.5 km), and strict unimodality, the overlap metrics are similar to the decoder. Finally, it is also empirically observed that in some cases (e.g. Fig. 4 d)) the model can underestimate the

400 peak of the spectrograms, which is a propagation of the decoder behavior. As a safety check, it was also verified that running
PAMTRA on the latent variables also led to spectrograms close to the original ones, as displayed on Fig. 4 c).

Let us point out at this stage that unlike most machine learning models, which are trained on a dedicated dataset and then
implemented on independent data, the encoder is here trained directly on the data of interest. Indeed, the aim is not to create
a generic model that can be used to retrieve microphysical variables from any dual-frequency spectrogram: the aim is to find

405 the latent variables which minimize the reconstruction error on specific measurements; the encoder can learn any relevant
feature from the input data to achieve this goal. In that sense, *overfitting* the data, which can be an issue in usual machine
learning problems, is not a concern when training the encoder. It is however preferable to train the model on a large enough
dataset rather than just a few spectrograms: this reduces the risk of converging toward local minima that would correspond to
non-physical combinations of microphysical parameters.
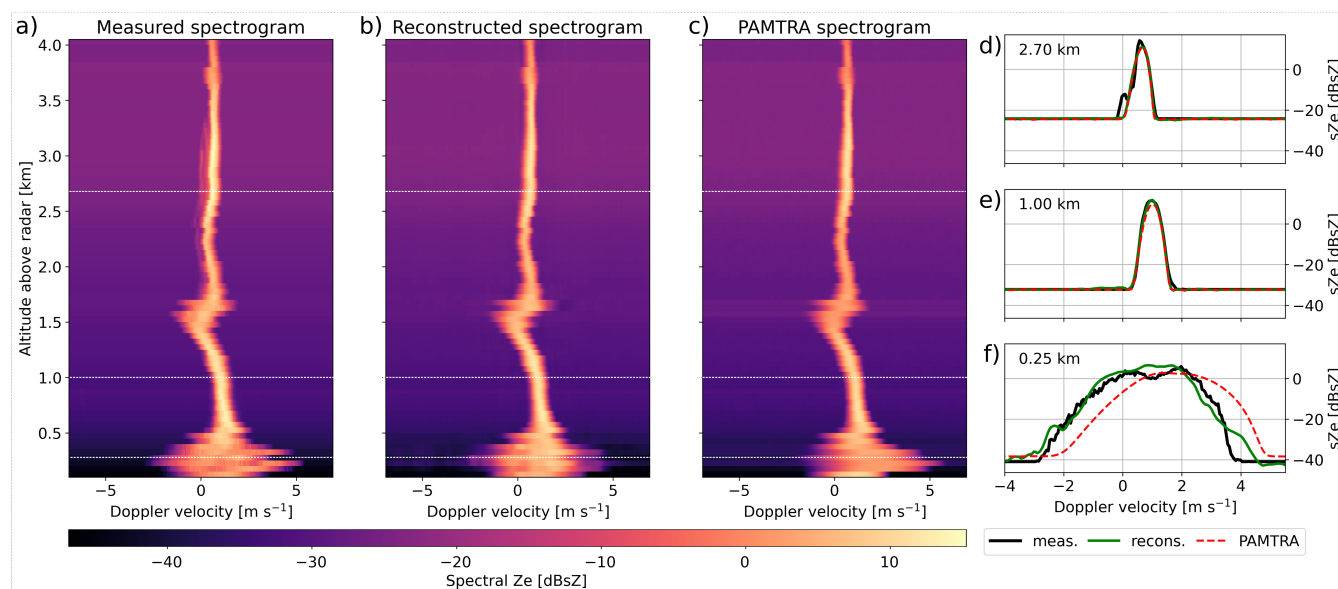
Figure 4. Examples of W-band spectrograms: a) measured, b) reconstructed through the pipeline, c) reconstructed in PAMTRA from the learnt latent features. Corresponding spectra from selected altitudes are displayed in panels d), e) and f); they are indicated with dashed white lines in the spectrograms. Units of spectral reflectivity are used, i.e. $1\,\mathrm{dBsZ} = 10\log_{10}(1\,\mathrm{mm}^6\mathrm{m}^{-3}(\mathrm{ms}^{-1})^{-1})$

.

410 ## 5.2 Qualitative assessment of the retrieval

5.2.1 Microphysical parameters

This section presents a qualitative perspective on the retrieval results, based on a snowfall event that took place on January 23, 2021, and through a visual inspection of the retrieved variables. Figure 5 features timeseries of some relevant radar measure-

ments (left column) and retrieved microphysical variables (right). The radar data include Ze_X, the dual-frequency reflectivity

415 ratio DFR_{XW}, Doppler velocity and spectral width (at W-band), as well as the hydrometeor classification from MXPol (cf Sect. 3). Let us highlight that the latter classification, derived from polarimetric variables, is completely independent from WProf and ROXI spectrograms. The microphysical variables included are IWC, D_0, b_m, β_a and A_r. The variables a_m and α_a, not shown, are highly correlated to respectively b_m and β_a (see e.g. Fig. 10 and B4).

A first general observation from the retrieval timeseries is the persistence of spatio-temporal structures visible in the radar

420 data, like the fall streaks. While the pipeline explicity took into account the *spatial* consistency of the measurements —through the use of convolutions—, the *temporal* features are never used in the training of the model. It is thus reassuring that the full spatiotemporal features are well captured by the retrieval method.

The retrieved values are also fairly consistent with the physical interpretation that stems from the radar measurements. IWC correlates quite strongly with Ze_X values, i.e. large IWC values are retrieved for strong reflectivity measurements (e.g.

425 around 15:10 and 15:50 UTC). The size parameter D_0 also matches the intuition, with small diameters near cloud top, and some localized pockets with large values e.g. around 15:10 between 1 and 2 km range, which correspond to regions of large dual-frequency ratio. The D_0 timeseries also agrees seemingly well with the hydrometeor classification that tends to identify aggregates in regions where D_0 is larger (with for example the same fall streak around 15:10 UTC).

430 The β_a exponent of the area-size relation features smaller values when Ze_X and DFR are low, which is compatible with small non disk-like particles such as columnar crystals, while larger values could indicate aggregates or rimed snowflakes.

Somewhat more noisy are the mass-size exponent b_m and the aspect ratio A_r, although their values and spatial trends still seem reasonable. They are rather correlated, which is not unrealistic: particles with aspect ratio near 1 are rounder and thus closer to spheres, which in turn, have a b_m close to 3. Indications of riming in the hydrometeor classification (visible

435 as yellowish-red regions) roughly correspond to regions with larger A_r and b_m, as expected from rimed particles (e.g. 15:10 between 0.5 and 1 km, 15:50 between 0 and 1.5 km, 16:00 around 1 km). Additionally, small values of b_m and A_r are retrieved near cloud top, consistent with crystal-like particles, while fall streaks where high D_0 values point to aggregation (15:00 to 15:20) also have medium-high b_m and A_r. A few time steps stand out with large values of A_r and b_m, coinciding with regions where large spectral width and variable v_{Dopp} suggest strong turbulence (e.g. 16:00, 1 km). In such high-turbulence cases, the

440 retrieval cannot be expected to perform perfectly since the shape of the spectra is then largely dominated by turbulent broadening.

Let us add a few words about correlations between certain variables. As mentioned before, some expected behaviors are observed in the retrieval like the apparent correlation between b_m and A_r, or between D_0 and β_a. This is not in any way

445 enforced by the pipeline, since those variables are prescribed independently when generating the training set. The correlations between a_m and b_m (and α_a and β_a), which are also expected, are slightly different: when building the training set (cf Sect. 3.1 and Appendix B1), these variables were sampled in a correlated way —with some noise included— to avoid completely

unrealistic combinations, and this may therefore influence the retrieval; however, the variables are not explicitly constrained during the training of the encoder: it is therefore reassuring to see that the model output still follows the expected correlations.
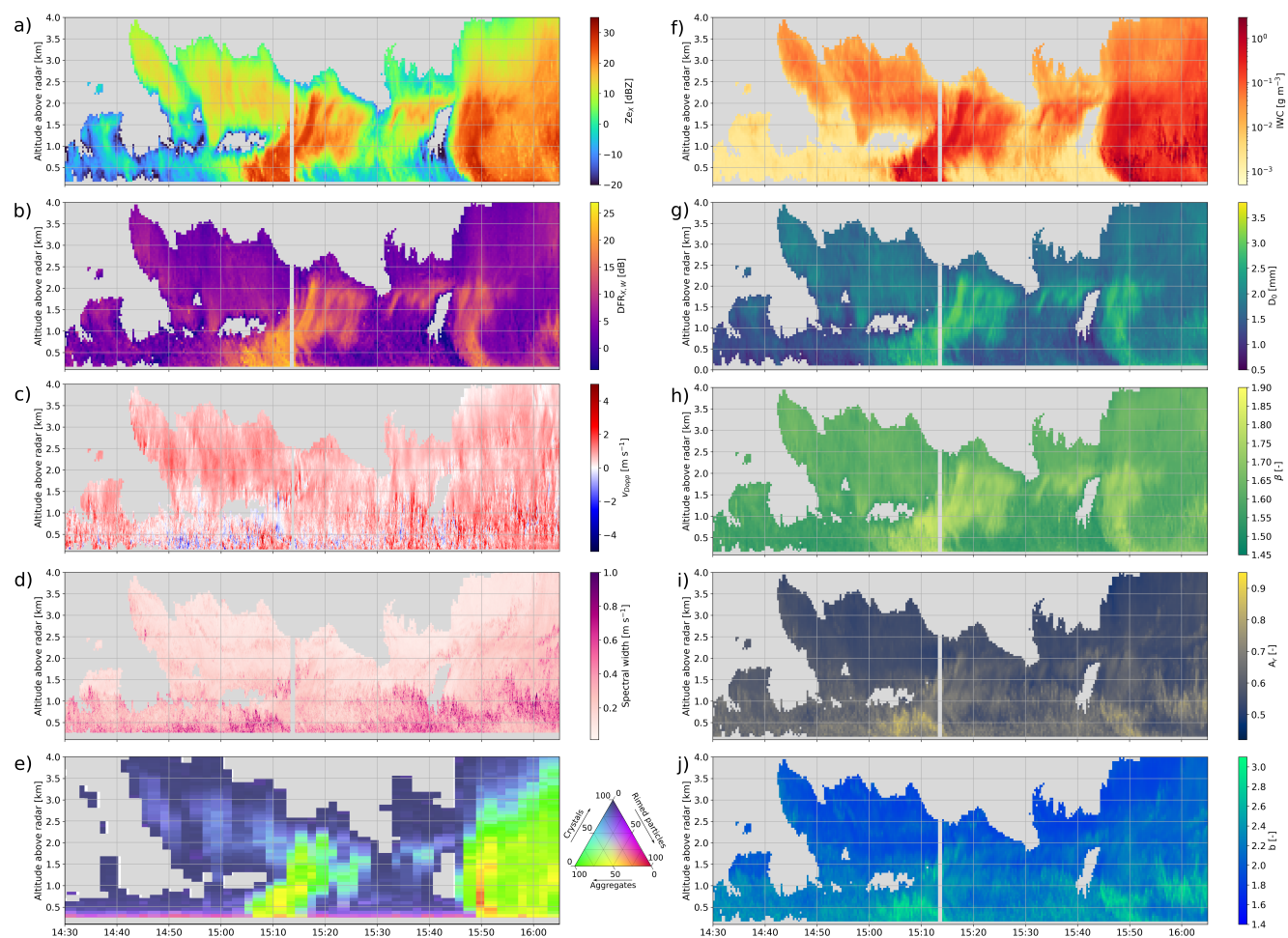
Figure 5. Height-time plots of radar measurements and microphysical retrievals. The left panels contain radar data: a) Ze_X, b) DFR, c) W-band mean Doppler velocity, d) W-band spectral width, e) Timeseries of hydrometeor classification with demixing showing the proportion of the three main particle types identified: here aggregates, rimed particles and crystals (cf. Section 3.3). The right panels feature microphysical retrievals: f) ice water content, g) size parameter D_0, h) area-size exponent β_a, i) aspect ratio A_r, j) mass-size exponent b.

5.2.2 Other retrieved variables

In addition to the microphysical descriptors, the latent features also comprise other quantities which are required by the pipeline in order to reconstruct the spectrograms (cf. Table 1). These are not designed to serve a proper physical interpretation, but their

behavior should still be assessed.

455   The noise level is only related to instrument properties and range gate, and in no way to microphysical or atmospheric processes. As visible in panels a) and b) of Fig. 6, the noise level estimates are exactly what could be expected and reflect the evolution of the radars' sensitivity with range. The slight temporal variations of the baseline level visible at X-band are due to the non-constant calibration issues of ROXI, mentioned in Sect. 3.2. At W-band, the abrupt change of sensitivity around 900 m range is due to the change of chirp table. Some artifacts are also visible (e.g. panel b), 16:00 at 1 km), in the same regions

460 of Fig. 5 where the other retrieved values visually also appeared less reliable. This is possibly related to the presence of strong turbulence in these areas —as suggested by enhanced spectral width and varying mean Doppler velocity—, which can indeed be expected to affect the retrieval's accuracy.

    The radial wind estimates serve to artificially correct for shifts of the Doppler spectra caused either by vertical wind i.e.

465 up- or downdrafts, by horizontal wind contamination by horizontal wind in the case of imperfectly vertical radar beams, or by biases in the velocity-size relation of the forward model. Their interpretation as a physical atmospheric quantity should thus be avoided. However, it is rather reassuring to see in panels c) and d) of Fig 6 that the X- and W-band are not too far off and especially that their cofluctuation is satisfactory: the opposite would be a problem since the Doppler velocity timeseries of both radars are rather similar (not shown). Similarly, the broadening parameters are similar in X- and W-band, and also somewhat

470 follow the spectral width (Fig. 5 d). We recall that the broadening parameters are not expressed in physical units, but as the size of the Gaussian kernel that results in the observed broadening. This is kept as such to highlight that these variables include all the broadening causes (i.e. not just turbulence but possibly also horizontal wind, etc.) and are rather a side-product of our retrieval than descriptors of actual atmospheric dynamics.

5.3   Comparison to in-situ data

475 In this section, we take a step further in the evaluation of the retrieval by performing quantitative comparisons with airborne in-situ measurements.

5.3.1   Ice water content

Figure 7 illustrates retrieval results of ice water content in comparison with in-situ estimates, computed as $IWC = TWC - LWC$ (cf. Sect. 3.3). In Fig. 7 a) and b) are displayed the timeseries of ice water content, first as a time-height plot to which

480 the aircraft trajectory is added, then along the aircraft trajectory — to which retrieval outputs are overlaid at timesteps of overpasses. The comparison is overall good, with satisfactory cofluctuations as well as reasonable agreement in the values themselves. For reference is also displayed the IWC retrieved from RASTA measurements (Delanoë et al., 2007, , c.f. Section 3.3)), which appears slightly more variable. In Figure 8 a), the scatterplot of retrieved to measured IWC combines the results from the three flights; the points are color-coded with Ze_X to illustrate that large IWC corresponds to large reflectivity, as

485 expected and already noted in the qualitative analysis. This scatterplot confirms the robustness of the retrieval results and their
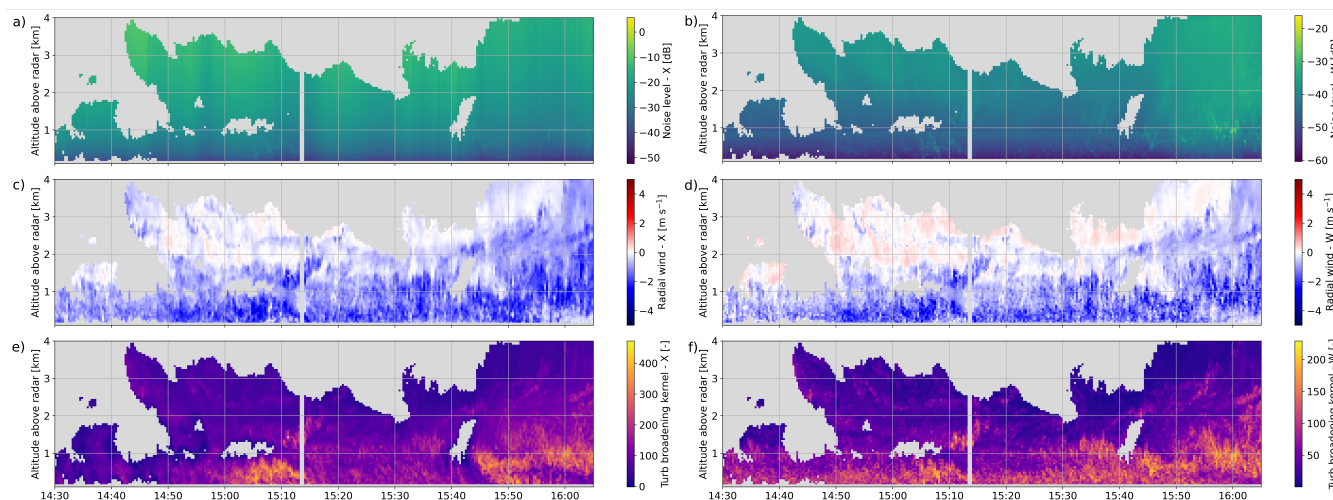
Figure 6. Time-height plot of additional retrieved variables: a) (resp. b)) Noise level at X- (resp W-) band; c) (resp. d)) radial wind at X- (resp. W-)band; e) Broadening at X- (resp W-)band

good correlation to the measured IWC (R = 0.85 in logarithmic scale), with however the existence of a slight bias toward low values (-0.36 in logarithmic scale). Surely, the spread of the values remains substantial, sometimes within orders of magnitudes, but it should be kept in mind that even at times of overpasses the aircraft is not perfectly colocated with the radar measurements, and that the sampled volumes are not identical; additionally, the single-frequency RASTA retrieval seems to have an even larger

490 variability than our retrieval.

5.3.2 Size parameter D_0

Aircraft measurements do not provide a variable that can directly be compared to the D_0 retrieved through our method. In order to obtain a comparable quantity, an exponential function was fitted to the aircraft PSDs; only diameters greater than $800\mu m$ were taken into account, since it was empirically noted that smaller particles — which have fainter signatures in radar data —

495 did not follow this exponential behavior. From each fit, a value of D_0 is derived, as shown in the timeseries of Fig. 9 a); in order to monitor the validity of this approach, the correlation coefficient of the fits are also included in the timeseries, and are typically very high (often R > 0.9). Our retrieval is superimposed to the timeseries, and compared to the in-situ values in Fig. 8 b) using all available flights. While this was not perceptible in the qualitative analysis, D_0 retrievals actually show a strong bias (+1.25 mm) when compared to aircraft measurements, leading to an overestimation of particle size. An investigation of

500 possible causes for this behavior is proposed in Section 6. This being noted, the cofluctuation between retrieved and in-situ D_0 is nonetheless good (R = 0.77), which gives confidence that the retrieval is still highly relevant for process-oriented studies: there, even more than the actual values, the changes and evolution of particle size can indicate the occurrence of specific snowfall growth or decay mechanisms.
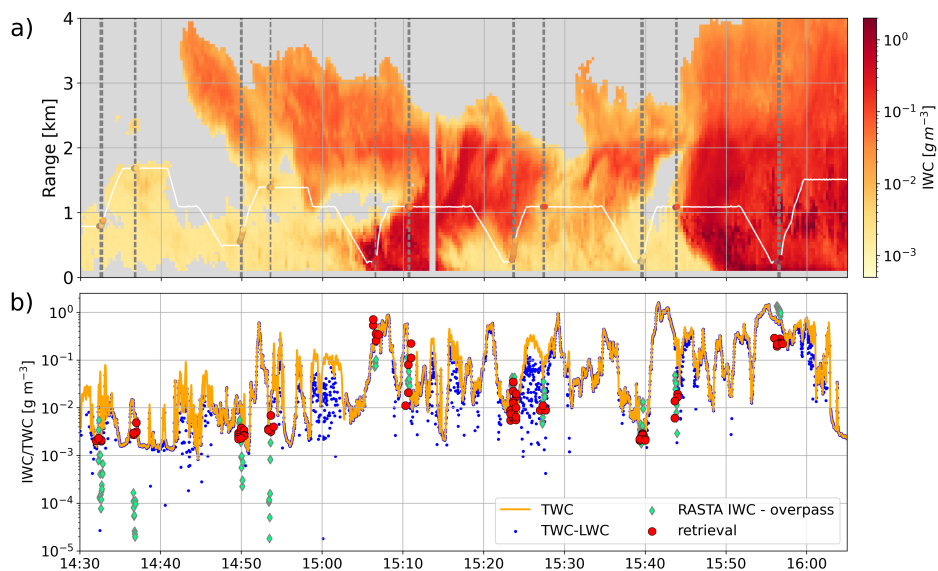
Figure 7. a) Time-height plot of IWC retrieval, to which the aircraft trajectory is overlaid as altitude as a function of time; aircraft IWC values at timesteps of aircraft overpasses (horizontal distance smaller than 1 km) are shown as scattered points. Dashed vertical lines indicate when the aircraft is within 500 m of horizontal distance to the radars. b) Timeseries of water content measured by the aircraft (TWC and $TWC - LWC$) and overlaid radar retrieval.

5.3.3 Mass- and area-size relations

505 Mass- and area-size power law coefficients are explicitly computed from the aircraft measurements and can therefore be compared to our retrieval. However, the timeseries of these aircraft quantities are highly noisy and thus point-to-point comparisons did not appear meaningful; it was therefore preferred to perform a statistical analysis. For each flight, we compare the histogram of b_m (resp. β_a) sampled by the aircraft during its entire flight (except for the part of the flight to and from the campaign location), to the histogram of retrieved b_m (resp β_a) above the radars, during the time frame of the flight and in the altitude

510 range sample by the aircraft, which excludes for instance regions near cloud top. The histograms of b_m agree rather well (Fig. 10 a)), with a similar mode around 2.2, although fewer values below 2 are retrieved. In Fig. 10 b), the histograms again have relatively close peak values (around 1.6 for the retrieval and 1.7 for the aircraft). There however, and for b_m to a lesser extent, the retrieved values histogram is much narrower than the aircraft one. This not too surprising, given how noisy the aircraft measurements are, and considering that the volume sampled by the PIP and 2DS probes is much smaller than the radar volume

515 —which automatically increases the variability and flattens the distribution. With this in mind, these histograms support a rather good consistency of the retrieval with the aircraft measurements.

In addition, we verify that the relations between a and b (resp. α-β), retrieved and measured, are consistent: this is visible in Fig. 10 c) (resp. d)), where the scatterplots of a vs. b (resp. α vs. β) are overlaid. Although not perfect, the match is reasonable. Let us highlight that once again, retrieved β values are narrower, consistent with the histograms.
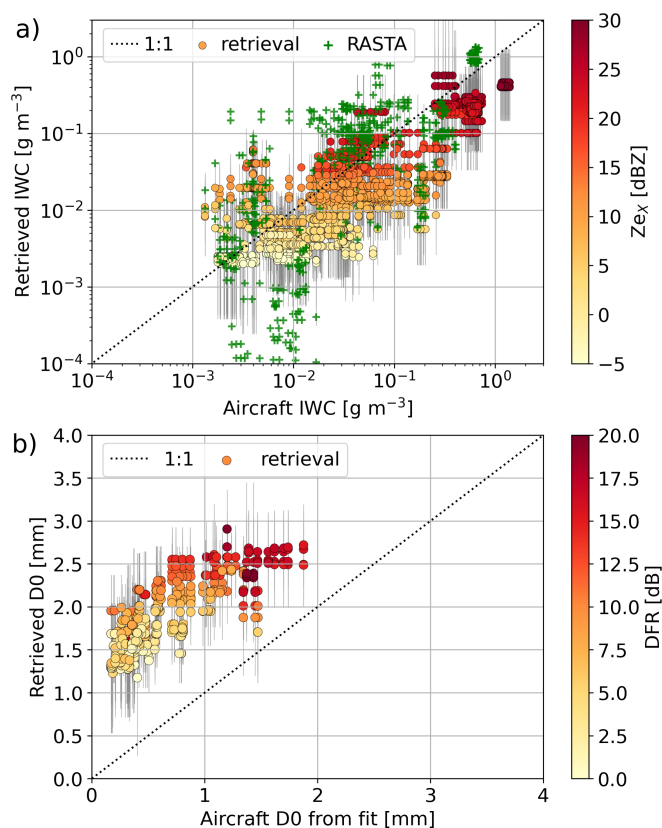
Figure 8. Scatter plots of retrieved vs. aircraft measurement of a) IWC and b) size parameter D_0. Each point corresponds to a time step when the aircraft is within 1 km horizontal distance to the radars. Three flights are used (Jan 22, Jan 23, Jan 27). Color indicates corresponding a) Ze_X and b) DFR. The black vertical lines indicate the standard deviation of the retrieval.

5.3.4 Aspect ratio

The last microphysical variable for which we can perform a comparison is the mean aspect ratio: similarly to the mass-size exponent, Figure 11 a) displays the histogram of retrieved and aircraft values. A significant difference is visible in the modes, with the aircraft values around 0.45 and the retrieval mode around 0.6; this however is coherent with the difference in the definitions of aspect ratio in each case. The aspect ratio retrieved through our pipeline is Ar_v defined as the ratio of particle dimension along vertical to maximum dimension, whereas the aircraft measurement is $Ar_\perp$ is the ratio of minor axis length to maximum dimension. Relating both quantities is not directly possible without having additional information on particle orientation, but an intuition can be gained from Fig. 11 b) where the relation between Ar_{vert} and $Ar_\perp$ is shown for particles randomly oriented within a certain angle (90° corresponds to completely random orientation). Using the relations of Fig. 11 b), a transformed histogram is included into panel a) showing the equivalent aircraft Ar_v assuming ellipsoidal particles with
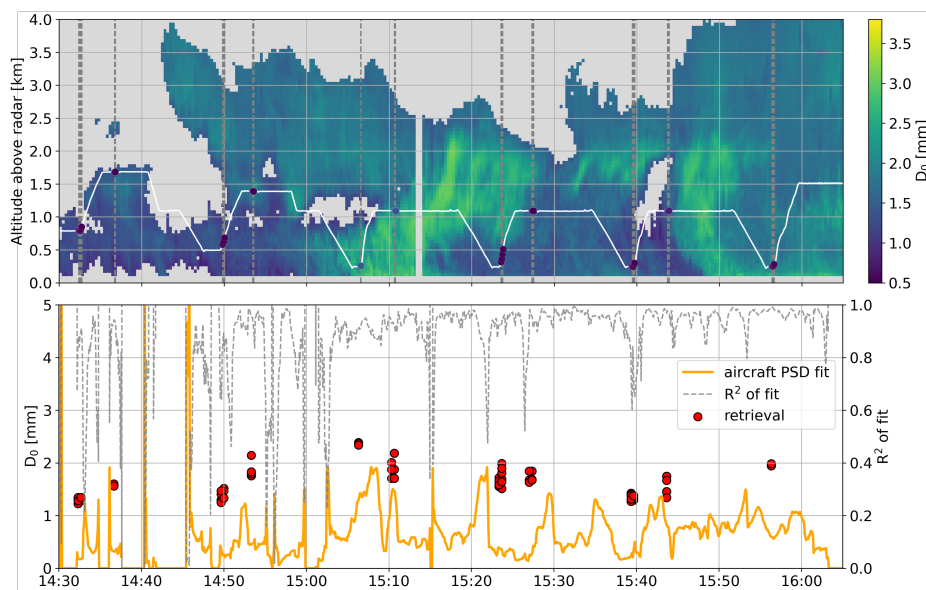
Figure 9. As for 7, but for D_0.

530   random orientation within $75°$: it fits rather well with the retrieval. While this is not per se rigorous, it gives a qualitative understanding of the observed discrepancy.

6   Discussion

While the results are overall encouraging, the previous section highlighted some points that call for further discussion. This section investigates the sensitivity of the pipeline to certain key hypotheses and provides some insight into possible causes for
535   the bias in D_0.

6.1   Sensitivity to calibration

One limitation of our framework is that it requires a good calibration of the radars, both absolute and relative, which is undeniably a crux of most radar-based retrievals. As detailed in Appendix A, in this case W-band reflectivity was assumed to be the ground truth, and then used to cross-calibrate the X-band radar. Although W-band reflectivity was corrected for
540   gaseous attenuation, we neglect attenuation in snowfall, which is undoubtedly an oversimplification. This can be particularly problematic if wet snow or supercooled liquid water is present, that can strongly attenuate millimeter-wavelength signal (with e.g. path-integrated attenuation up to 5 dB for liquid water paths of $500\ \mathrm{g\ m^{-2}}$, Kneifel et al., 2015).

In Fig. 12 are displayed the biases in retrieved IWC and D_0, computed as the mean difference between retrieved values and aircraft measurements, when a constant offset in reflectivity is added to the input X- and W-band spectrograms. The following
545   behavior is observed, in accordance with previous qualitative observations: IWC is especially sensitive to X-band reflectivity
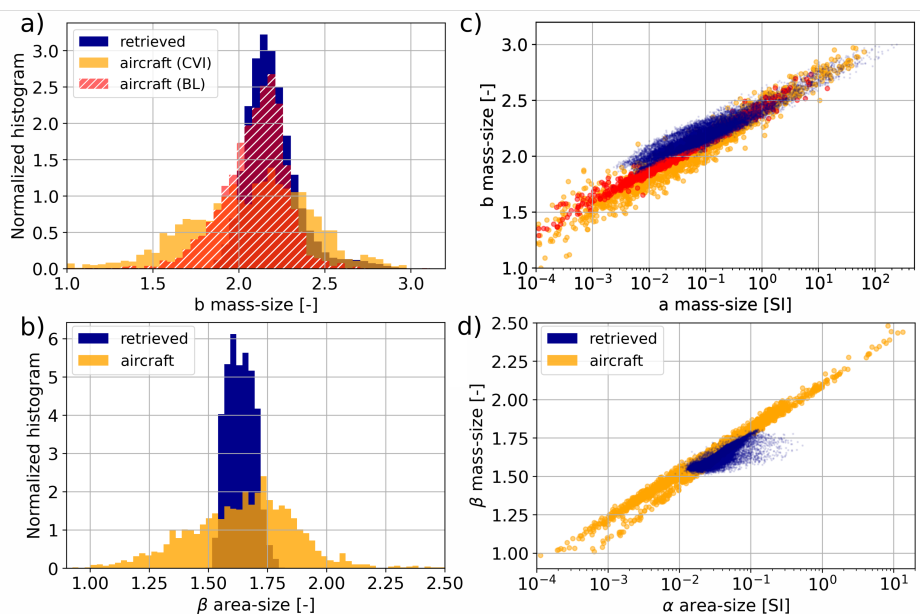
Figure 10. Histograms of a) mass-size and b) area-size exponents. Scatter plots of c) mass-size and d) area-size exponent to prefactor, from retrieval and aircraft measurements (cf. Sect. 3.3)
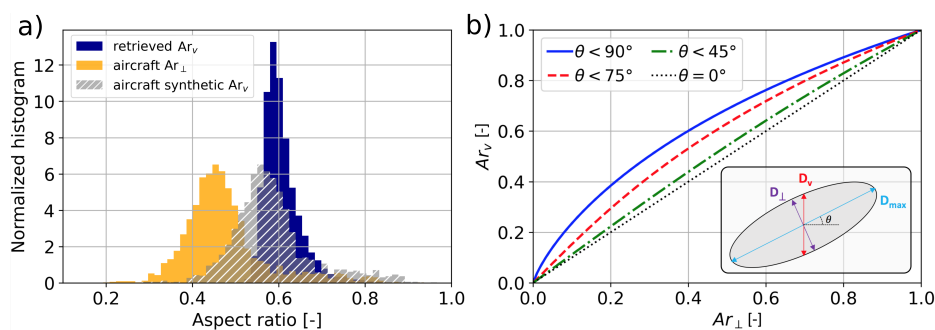
Figure 11. a) Histogram of retrieved and aicraft-measured aspect ratio. b) Illustration of the relation between $A_{r,\perp}$ and $A_{r,v}$ for particles with random orientation within a given angle θ; the various quantities are sketched in the bottom right of panel b).

(as illustrated in panel b and by the rather horizontal color gradient in panel a). Rather, D_0 is more sensitive to differential offset i.e. to changes in the DFR (as illustrated in panel d and by the rather diagonal color gradient in panel c).

While this Ze calibration is undoubtedly a key factor in the uncertainty of the algorithm, it does not appear to cause extreme divergence in the retrieval: the changes in D_0 and IWC shown in Fig. 12, while not negligible, are also not massive. In

550 particular, Ze miscalibration could not explain solely the observed D_0 discrepancy: changing the DFR of $-6dB$ —which is substantial— only brings down the bias from $1.25\,\mathrm{mm}$ to around $0.9\,\mathrm{mm}$.

Let us mention that possible improvements of the framework could overcome this requirement of perfect reflectivity calibration, by incorporating attenuation as an additional retrieved parameter; this would however increase the ill-posedness of the problem, and was therefore not implemented in these first stages of development.
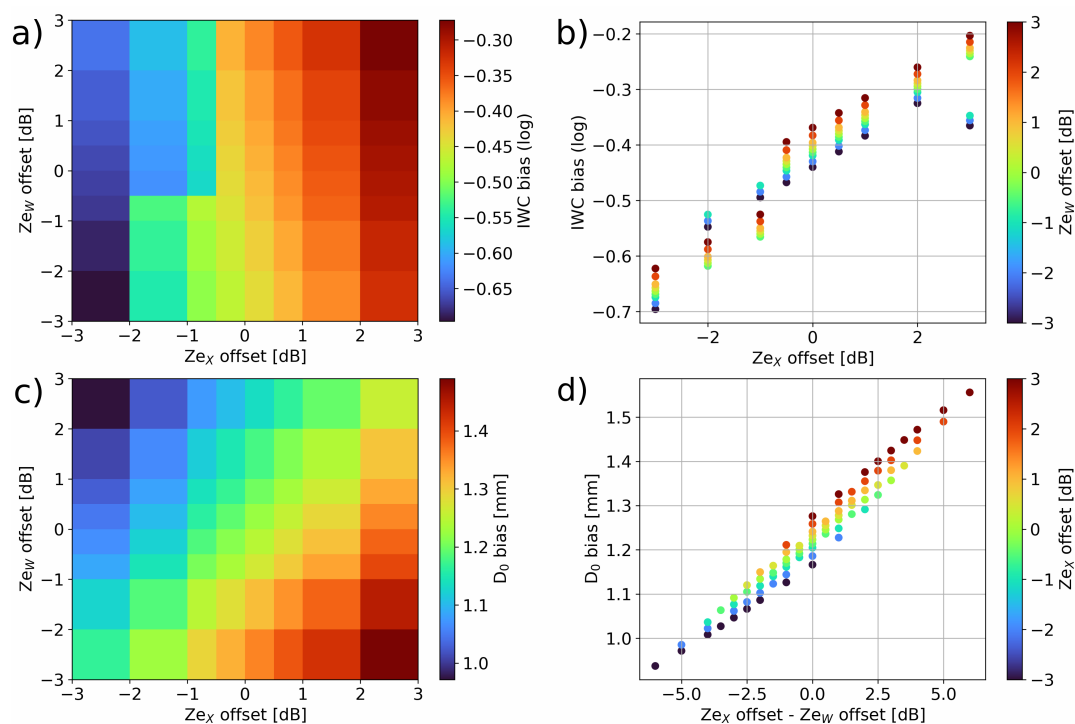
Figure 12. a) Heat map of IWC bias as a function of X- and W-band reflectivity offset; the bias is computed using the aircraft values as reference b) Different visualization showing IWC bias as a function of Ze_X offset. c) As for a), but for D_0. d) As for b), but for D_0 and offset in dual-frequency ratio.

6.2 Training set limitations

Another aspect of our framework which could cause a bias in the retrieval is if the training set is too narrow. While special attention was paid to this potential issue as the microphysical parameters were sampled from the MASC database, there is likely still room for improvement. In particular, the size cutoff for good-quality images in the MASC is quite high and very few particles with a diameter below 0.5 mm are accurately captured. For reference, Fig. 13 illustrates the histogram of D_0

560 derived from MASC measurements (Grazioli et al., 2022), and by the aircraft 2D-S and PIP probes during the ICE GENESIS campaign. The limitation is apparent: the aircraft is able to capture much smaller particles, but not beyond a certain size, while the MASC can detect large snowflakes but very few small particles. The framework would thus probably benefit from training

the decoder on a larger dataset, that would include a better representation of this smaller particle range. It is yet unlikely that this would entirely resolve the size bias, for there is still an overlap between the aircraft-measured size range and that on which the model was trained.
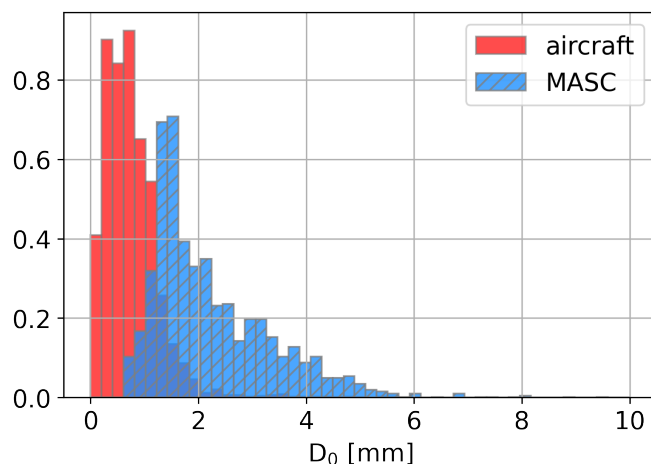
Figure 13. Histogram of D_0 from aircraft measurements (PIP and 2D-S) during ICE GENESIS (red) and from MASC measurements (MASCDB)

565

6.3 Scattering model

Surely one of the strongest hypotheses on which the pipeline was built is the parameterization of the scattering model in forward simulations. As explained in Sect. 3.1.1 and Appendix B2, the default version of PAMTRA was used which to this date (28/06/2021) assumes constant values for certain parameters of the SSRGA, and allows to change two coefficients (κ and

570 β, cf. Hogan and Westbrook (2014) and online documentation (PAMTRA team, 2015)). Several studies suggest however that more parameters are needed, and moreover that their values can vary significantly (e.g. Leinonen et al., 2018a; Ori et al., 2021) depending on particle type, shape, etc.

To get an empirical sense of how this could affect the retrieval, the approach described hereafter was followed. First, a few time and range gates were randomly selected from the dataset. The corresponding retrieved values were then modified by

575 adding a D_0 offset ranging from -1.5 to +2 mm, and PAMTRA simulations were run on the microphysical parameters obtained, using the same settings as in Sect. 3.1.1. Parallel to that, slight modifications of the PAMTRA code were conducted to allow for modification of the four literature coefficients of the self-similar Rayleigh-Gans approximation (κ, β, γ, ζ); new simulations were run for the selected time and range gates, keeping the retrieved microphysics unchanged but randomly changing the SSRGA parameters within $\pm 10\%$ of their original values. As seen in Appendix B2, this is well within the typical variability

580 of the coefficients calculated from simulating various types of particles (e.g. Leinonen et al., 2018a, , Fig. 5). It was then verified by visual inspection that changes in D_0 and changes in SSRGA coefficients did not result in very different spectrum

shapes but mostly affected the amplitude of the spectra: hence, the influence of the changes were measured by the change in the scalar "total" reflectivity at X- and W-band, then in the dual-frequency ratio. The obtained results are illustrated in Fig. 14 (detail in the caption): they suggest that moderate changes in the SSRGA parameters could have an impact similar as 585 varying the size parameter of approximately 1 mm (-0.6 to +1.5 mm), which is a significant change. Taking this investigation a step further, the influence of each of the four parameters can be computed independently, by following the same steps but changing only one coefficient: it appears that the output is most sensitive to β and γ, which each cause amplitude changes corresponding to ΔD_0 of at least $\pm 0.4 mm$. Obviously, this empirical analysis cannot be directly translated into a quantitative interpretation; yet it highlights that the scattering model can have a substantial influence on the retrieval. This leads us to 590 believe that the D_0 bias observed when comparing our retrieval to aircraft measurements is partly caused by an inaccurate or insufficient parameterization of the radiative transfer model. In order to remedy this effect, a forward model with a more subtle parameterization is likely required when designing the decoder training set.
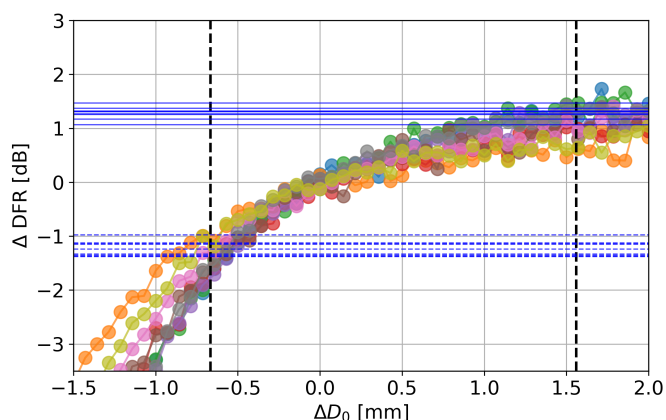
Figure 14. Colored lines with scattered points: $\Delta SDFR$ caused by adding a diameter offset ΔD_0 on microphysical descriptors of selected (time, range) gates. Horizontal lines: for each of these (time,range) gate, maximum ΔDFR (positive and negative) caused by a modification of the SSRGA coefficients within $\pm 10\%$. For each selected (time,range) gate, the intersection of the horizontal and colored lines give a ΔD_0 value which causes the same change in DFR as a change in SSRGA coefficients (worst case). Dashed vertical lines show the mean of these ΔD_0 values.

Another point should be briefly mentioned regarding small particles, which are Rayleigh scatterers at both X- and W-band. This means that if a population is composed entirely of small particles, the influence of particle size and number concentration 595 is hardly distinguishable in the spectrograms. The ill-posedness of the problem is reinforced and the retrieval could be expected to have a reduced accuracy, even if the training set and scattering model were improved.

6.4 Comparison to other frameworks

In this section, we discuss more broadly the pipeline that was developed, in comparison with other possible approaches. As argued in Sect. 2, we believe that the novel framework introduced here is a key aspect of this work.

600 In order to support this point, a *direct* deep learning inversion model was also designed: it essentially consists in learning the inverse of our decoder, similar to the approach of e.g. Chase et al. (2021). It is presented in Appendix C and the results show that although still respectable, this direct retrieval is noisier and less accurate than our model due to its ill-posedness. For instance, when comparing retrieved values to in-situ measurements of IWC, the correlation coefficient drops from R = 0.85 (with the proposed method) to R = 0.59 (with this direct inversion). Similar behaviors are observed for the other retrieved

605 variables.

Let us mention an alternative approach that could be used, which lies half-way between classical OE and the proposed method. The notations used here are those of Sect. 2. Once a differentiable approximation of the forward model is known ($\tilde{\mathbf{f}}$), another way to look for X_r is to find the minimizing argument of $||\tilde{\mathbf{F}}(X) - Y_r||^2$ using gradient descent; a regularizing

610 term can be added to ensure for instance the spatial continuity of X or to enforce some degree of spatio-temporal smoothness. This requires only one deep learning model instead of two, and could thus seem more appealing, but the first approach was preferred. Indeed, by actually learning an approximation of the inverse mapping $\tilde{\mathbf{G}}$, and doing so on a large dataset, the risk of reaching a local minimum in X is reduced. Our method also does not rely any prior assumption on X, nor on any property of the latent space, like spatial smoothness; rather, it is constrained by the spatial structure of the observed signal itself.

615 ## 7 Conclusion

In this work, we proposed a new method for the retrieval of seven microphysical properties of snowfall from dual-frequency Doppler radar spectrograms. To our knowledge, no previous method allowed for the joint retrieval of these descriptors, and with this high spatial and temporal resolution. Some typical challenges of Doppler spectral retrievals were overcome, like the need for perfectly vertical beam alignment, or the requirement of very low turbulence, thus allowing for microphysical

620 retrievals in a larger range of atmospheric conditions. The approach relies on a two-step deep learning framework: a *decoder* network serves as a differentiable gate-to-gate emulator of a known radiative transfer model, while the *encoder* network learns to map the Doppler spectrograms to full profiles of microphysical variables. The algorithm could be assessed thoroughly by confronting the retrieved quantities to in-situ aircraft measurements which were conducted during the 2021 ICE GENESIS campaign. Overall, the comparisons with in-situ data are highly encouraging and support the validity of the framework: good

625 co-fluctuations as well as similar statistics are reported. Certain discrepancies were nonetheless observed: in particular, the retrieved values of the size parameters are affected by a bias, for which possible explanations were proposed. They point to limitations in the training set itself (in which small particles are under-represented), and to assumptions in the scattering model (which relies on the SSRGA). This analysis opens up for possible improvements of the method, particularly along the line of radiative transfer modeling. Meanwhile, in spite of these limitations, the method can provide relevant insights into snowfall

630 properties in the perspective of process-oriented studies whose focus is typically the relative spatial and temporal evolution of microphysical variables, rather than their exact numerical values. The approach could potentially be extended to include other variables and further alleviate the baseline assumptions, e.g. by using more generic PSDs or even retrieving scattering

coefficients, although computational constraints should be kept in mind. It should be kept in mind that the addition of new parameters increases further the ill-posedness and that two-frequency Doppler spectrograms may not be able to resolve it.

635   Given the convincing results obtained in other studies (e.g Mroz et al., 2021) with triple-frequency data, it is likely that our method would also gain in robustness and precision with the inclusion of an additional frequency.

The theoretical pipeline itself is an important contribution of this work, for it can be implemented in other settings and for different types of inverse problems. One fundamental difficulty of such problems is often their ill-posedness: several combinations of physical parameters can yield similar observations. The proposed approach mitigates this by learning information

640   from the spatial structure of the data thanks to convolutional neural networks.

Code availability.   The code for generating the training set using PAMTRA, and for training the decoder and encoder models, is available at https://github.com/annecbroux/DeepSpectralRetrieval (latest version).

Data availability.   The data of the ICE GENESIS campaign will be made available on the *Aeris* platform (https://en.aeris-data.fr/). The entire dataset will be made fully public, after an embargo period and the necessary post-processing time, at the beginning of 2023. Individual access

645   can be granted before this time upon contact with the authors.

Appendix A: Note on radar calibration and attenuation issues

In order to correct for the reflectivity offset between the two frequencies caused by instrument miscalibration, a relative cross-calibration was performed. The main issue encountered during this process was that ROXI's calibration was found to be time variable, for a reason that is not fully clarified yet —possibly related to a hardware artifact causing either the output power

650   or received secondary wave trains to fluctuate; the investigation of this issue is beyond the scope of this work. WProf was calibrated only a few days prior to the deployment, and an automatic recalibration routine (Küchler et al., 2017) was performed every 15 minutes, which mitigates calibration drift. Gaseous attenuation at W-band was corrected for using ERA5 atmospheric profiles and PAMTRA computations. Because of the ROXI-related issues and the accurate calibration of WProf, it was decided to use the latter radar reflectivity as a reference during snowfall events. The cross-calibration is then performed by matching

655   the X-band to the W-band reflectivity in regions were Rayleigh scattering is expected, chosen as the region with the lowest reflectivities close to cloud top. This relies on the relatively strong assumption that W-band attenuation in snowfall can be neglected; the presence of supercooled liquid water is also a potentially important source of attenuation which is not accounted for. This should be considered as a limitation of our method at its current stage, and motivated the discussion of 6.1 about the sensitivity of our model to reflectivity offsets.

660

A further point should be added regarding one event in the data set (January 22) which featured rain at ground level during a few hours before a transition to snowfall in the evening. Here, the calibration of ROXI was corrected with a constant offset

computed using the hours of the event with snowfall at ground level; Ze_W during rain was then corrected using Ze_X as a reference, for W-band attenuation in rain is evidently no longer negligible.

Appendix B: Training set

B1   Distributions

Fig. B1 illustrates the distributions of microphysical parameters in the MASC database (Grazioli et al., 2022) from which the decoder training set is sampled. The particles were grouped into four large categories: aggregates, planar crystals, columnar crystals, and graupel, using the classification output of Praz et al. (2017). To generate the training set, D_0, b_m and β_a are sampled using skewnorm fits of these distributions; a_m and α_a are sampled using their relation to b_m and β_a, as illustrated in Fig. B3 and Fig. B4. The histograms of other variables in the decoder training set are in Fig. B2.

B2   Self-Similar Rayleigh-Gans Approximation

The coefficients of the SSRGA made available by Ori et al. (2021) through the SnowScatt model were grouped by types of particles in order to match the four main categories used to sample the training set: planar crystals, columnar crystals, aggregates, graupel. The SSRGA in general and here the simulations of Ori et al. (2021) are mostly targeted on aggregates, so it was decided to include e.g. columnar aggregates in the columnar category, with the reasoning that when the size regime is that of columnar crystals, the scattering properties would approach those of the individual particles. While this rationale could be debated, it would most likely not trigger diverging results since the SSRGA collapses to Rayleigh scattering for small particles, meaning the exact values of coefficients have little impact. After grouping the particles into the different categories, the coefficients κ and β are then averaged within each group, for each size bin, as shown with the black lines in Fig. B5. In addition to the limitations already discussed in this study, mostly focused on the use of only two parameters and on the reduction of complexity by averaging, an additional point can be noted. It is strictly speaking not valid to assume a single set of scattering coefficients in SSRGA computations for an entire particle size distribution, since the coefficients are size-dependent. However, as underlined in Ori et al. (2021), the coefficients do not change significantly for large particles, while for small particles, it was already noted that the SSRGA simplifies to Rayleigh scattering regardless of the coefficient values, which makes this assumption altogether reasonable.

Appendix C: Alternative deep learning method

One of the motivations for using the architecture proposed in this work is the ill-posedness of the problem, which is arguably an obstacle for direct inversion methods. Nonetheless, such an inversion was also implemented, through a deep learning framework trained on the same synthetic dataset as the one used to train the decoder. This time, the input consists of dual-frequency spectra, and the output is the set of microphysical and atmospheric descriptors (same as Table 1). The architecture, not detailed here, is virtually the same as the encoder (cf. Fig 2 b), except that 2-dimensional convolutions are now 1-D: the neural network is
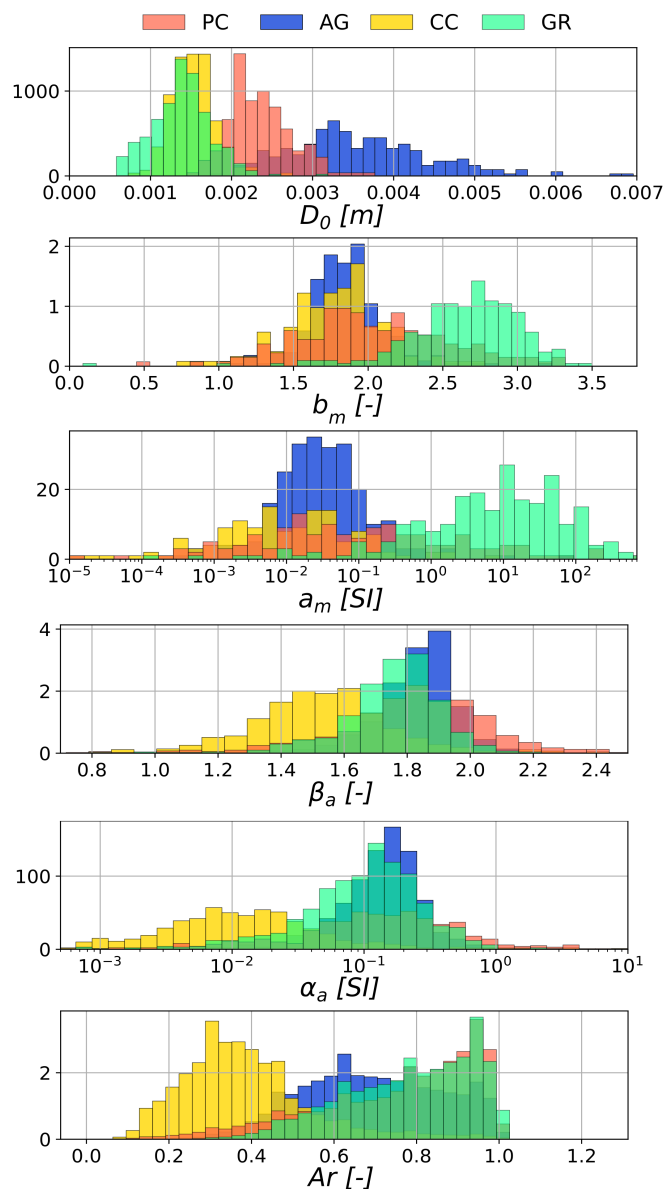
Figure B1. Histograms of the microphysical parameters in the MASC database. PC: planar crystals, AG: aggregates, CC: columnar crystals, GR: graupel

not trained on full spectrograms but on single-gate spectra, thus the range dimension is equal to 1. After training and tuning, the model is applied to the ICE GENESIS dataset. Figure C1 shows the same variables as in the left panels of Fig. 5, retrieved through this *direct* inversion. Overall, the order of magnitude of the variables is similar to that obtained with the new pipeline, and the very general spatio-temporal structure is also visible. This is reassuring since it suggests that the training dataset was
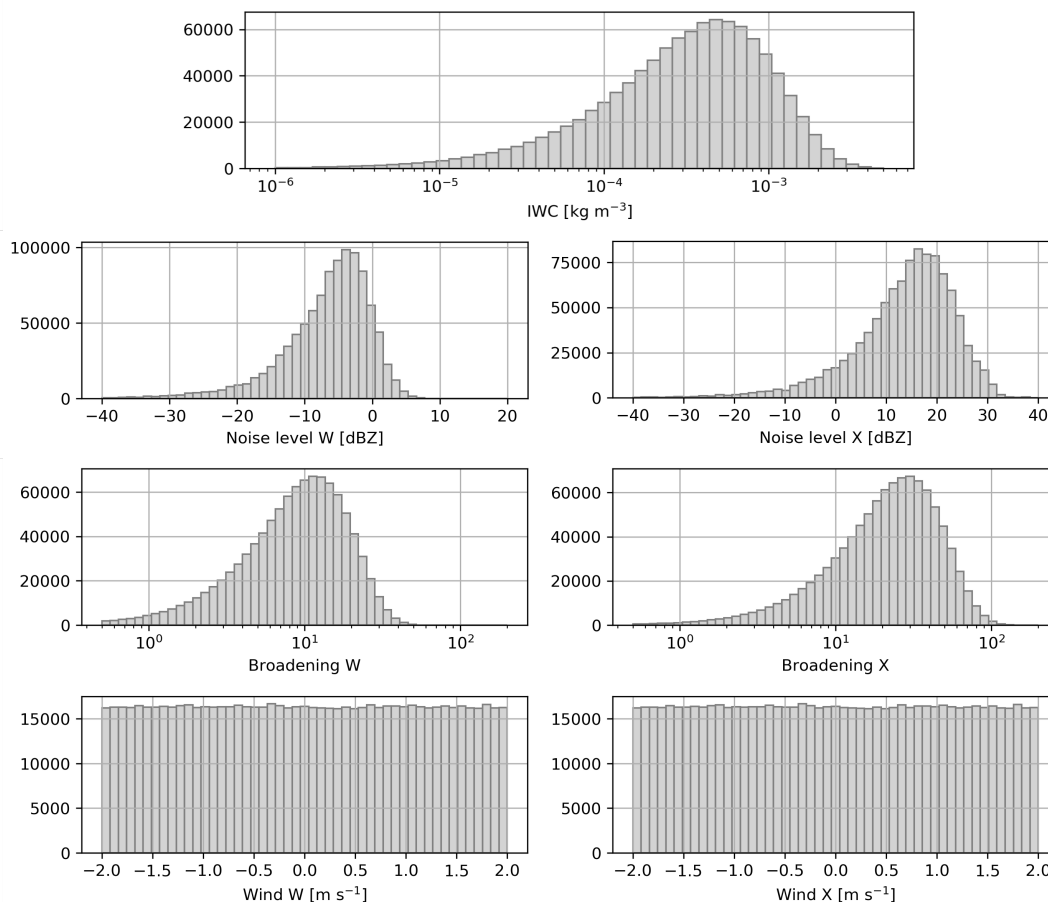
Figure B2. Same as B1 but for the other microphysical descriptors, not provided in MASCDB.

appropriate and indeed captured the scope of possibly observed spectra. However, it is also apparent that the retrieved variables are substantially noisier than through our method, reflecting the ill-posedness issue. When comparing these retrieved results with aircraft in-situ measurements, as done in Section 3.3, we obtain for example R = 0.59 for IWC (instead of R = 0.85).

700   Some variables also reach unrealistic values, e.g. aspect ratio close to 0 or even negative values of D_0.

Author contributions. ACBR and AB designed the study, with input from GG in the conception of the retrieval framework. ACBR implemented the deep learning pipeline with contributions from GG. The radar data were prepared by AM, NV and ACBR. Aircraft in-situ measurements were processed by LJ. Comparisons of retrieved to in-situ values, and sensitivity analyses, were conducted by ACBR with input from AB. ACBR prepared the manuscript with contributions from GG and AB and supervision from AB. All authors have read and

705   agreed to the published version of the paper.
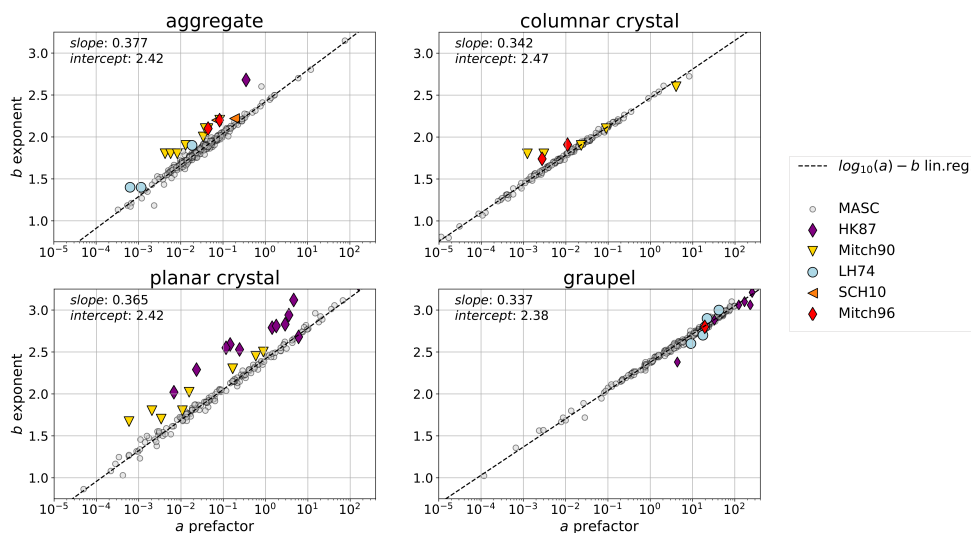
Figure B3. Relation between exponent and prefactor of mass-size relations for different particle types, computed using MASCDB.
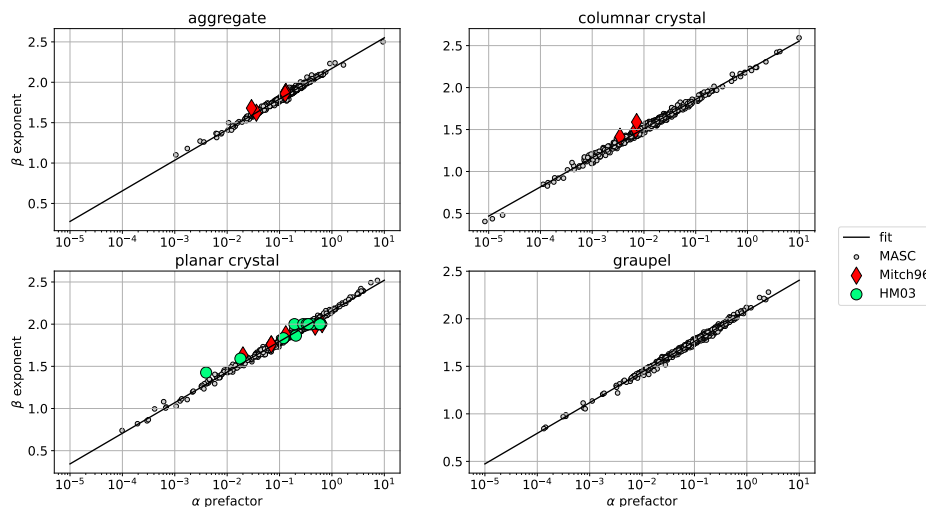
Figure B4. Same as B3 but for area-size relation.

Competing interests. At least one of the (co-)authors is a member of the editorial board of Atmospheric Measurement Techniques.

Acknowledgements. This project has received support from the European Union's Horizon 2020 research and innovation program under grant agreement No 824310 (ICE GENESIS project). Airborne data were obtained using the aircraft managed by Safire, the French facility for airborne research, an infrastructure of the French National Center for Scientific Research (CNRS), Météo-France and the French National
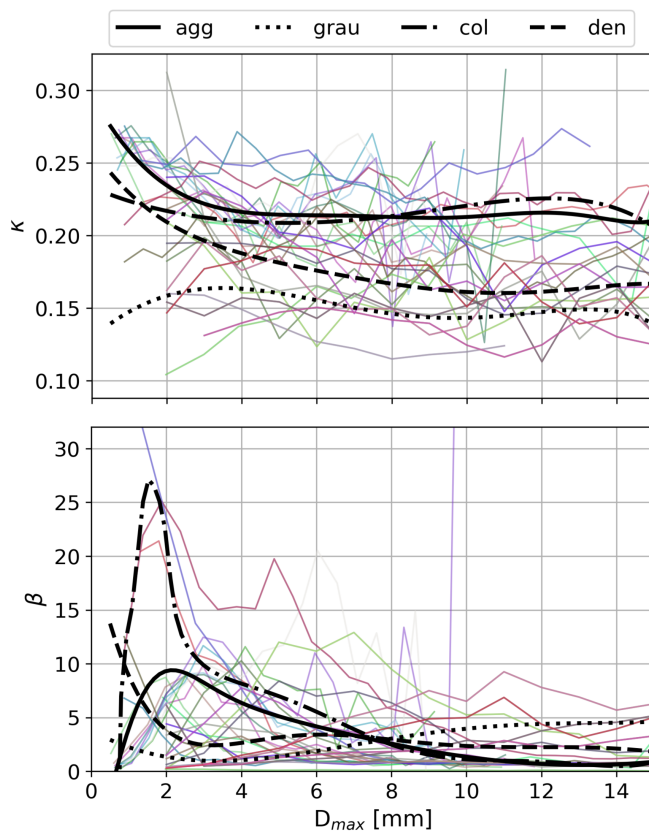
Figure B5. Colored lines: coefficient of the SSRGA computed in Ori et al. (2021) for a given particle type (as a function of D_{max}) for a) κ and b) β. Black lines: average coefficient when grouping by particle type, used for sampling the training set.

710   Center for Space Studies (CNES). Most of the microphysical in-situ data were collected using instruments from the French Airborne Measurement Platform, a facility partially funded by CNRS/INSU and CNES. We thank Davide Ori for his help in the initial parameterization of PAMTRA. Finally, we are grateful to Julien Delanoë and Susana Jorquera for providing the ice water content retrieved from airborne RASTA radar measurements.
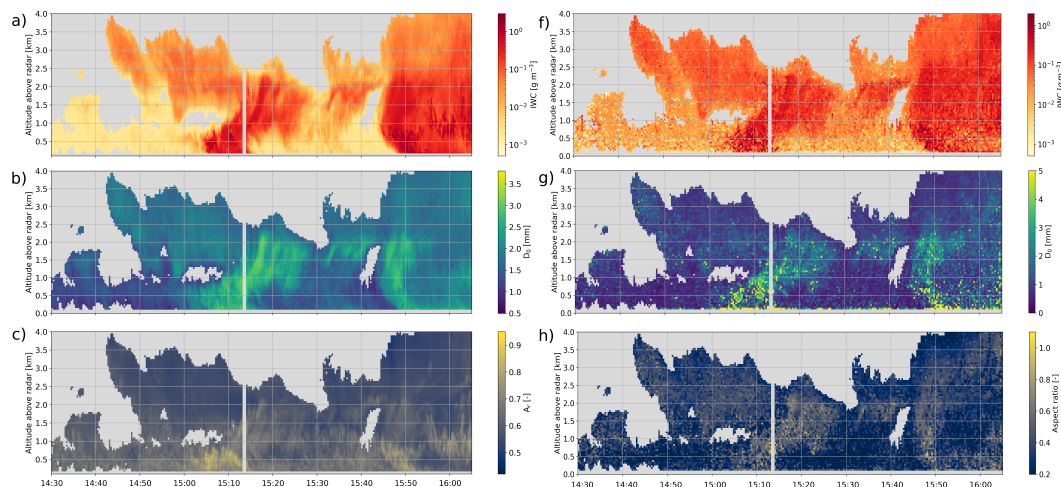
Figure C1. Comparison of timeseries for three examples of variables (IWC, D_0 and A_r) retrieved through the proposed framework (left panels) or a direct deep-learning retrieval (right panels). Note that the colorbars may differ (adjusted to reflect at best the variability in each field).

References

715   Anderson, T. L., Covert, D. S., and Charlson, R. J.: Cloud droplet number studies with a counterflow virtual impactor, Journal of Geophysical Research, 99, 8249–8256, https://doi.org/10.1029/93JD03522, 1994.

Barrett, A., Westbrook, C., Nicol, J., and Stein, T.: Rapid ice aggregation process revealed through triple-wavelength Doppler spectra radar analysis, Rapid ice aggregation process revealed through triple-wavelength Doppler spectra radar analysis, 19, 5753–5769, https://doi.org/10.5194/acp-2018-836, 2019.

720   Bauer, P., Dueben, P. D., Hoefler, T., Quintino, T., Schulthess, T. C., and Wedi, N. P.: The digital revolution of Earth-system science, Nature Computational Science, 1, https://doi.org/10.1038/s43588-021-00023-0, 2021.

Baumgardner, D., Abel, S. J., Axisa, D., Cotton, R., Crosier, J., Field, P., Gurganus, C., Heymsfield, A., Korolev, A., Krämer, M., Lawson, P., McFarquhar, G., Ulanowski, Z., and Um, J.: Cloud Ice Properties: In Situ Measurement Challenges, Meteorological Monographs, 58, 9.1–9.23, https://doi.org/10.1175/amsmonographs-d-16-0011.1, 2017.

725   Besic, N., FiguerasVentura, J., Grazioli, J., Gabella, M., Germann, U., and Berne, A.: Hydrometeor classification through statistical clustering of polarimetric radar measurements: A semi-supervised approach, Atmospheric Measurement Techniques, 9, 4425–4445, https://doi.org/10.5194/amt-9-4425-2016, 2016.

Besic, N., Gehring, J., Praz, C., Figueras i Ventura, J., Grazioli, J., Gabella, M., Germann, U., and Berne, A.: Unraveling hydrometeor mixtures in polarimetric radar measurements, Atmospheric Measurement Techniques, 11, 4847–4866, https://doi.org/10.5194/amt-11-
730   4847-2018, 2018.

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

Chantry, M., Christensen, H., Dueben, P., and Palmer, T.: Opportunities and challenges for machine learning in weather and climate modelling: Hard, medium and soft AI, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 379, https://doi.org/10.1098/rsta.2020.0083, 2021.

Chase, R. J., Nesbitt, S. W., and McFarquhar, G. M.: A dual-frequency radar retrieval of two parameters of the snowfall particle size distribution using a neural network, Journal of Applied Meteorology and Climatology, 60, 341–359, https://doi.org/10.1175/JAMC-D-20-0177.1, 2021.

Curry, J. A., Schramm, J. L., Rossow, W. B., and Randall, D.: Overview of Artic Cloud and Radiation Characteristics, Journal of Climate, 9, 1731–1764, https://doi.org/https://doi.org/10.1175/1520-0442(1996)009<1731:OOACAR>2.0.CO;2, 1996.

Delanoë, J., Protat, A., Bouniol, D., Heymsfield, A., Bansemer, A., and Brown, P.: The characterization of ice cloud properties from Doppler radar measurements, Journal of Applied Meteorology and Climatology, 46, 1682–1698, https://doi.org/10.1175/JAM2543.1, 2007.

Doviak, R. J. and Zrnic, D. S.: Doppler Radar and Weather Observations, https://doi.org/10.1016/C2009-0-22358-0, 1993.

Draine, B. T. and Flatau, P. J.: Discrete-Dipole Approximation For Scattering Calculations, Journal of the Optical Society of America A, 11, 1491, https://doi.org/10.1364/JOSAA.11.001491, 1994.

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

Geer, A. J.: Learning earth system models from observations: machine learning or data assimilation?, Philosophical Transactions of the Royal Society A, 379, https://doi.org/10.1098/rsta.2020.0089, 2021.

Geng, Z., Yan, H., Zhang, J., and Zhu, D.: Deep-Learning for Radar: A Survey, IEEE Access, 9, 141 800–141 818, https://doi.org/10.1109/ACCESS.2021.3119561, 2021.

Glorot, X. and Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, edited by Teh, Y. W. and Titterington, M., vol. 9 of *Proceedings of Machine Learning Research*, pp. 249–256, PMLR, Chia Laguna Resort, Sardinia, Italy, https://proceedings.mlr.press/v9/glorot10a.html, 2010.

Grazioli, J., Ghiggi, G., Billault-Roux, A.-C., and Berne, A.: MASCDB, a database of images, descriptors and microphysical properties of individual snowflakes in free fall, Scientific Data, 9, 186, https://doi.org/10.1038/s41597-022-01269-7, 2022.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1951–1954, 2015.

Heymsfield, A. J. and Westbrook, C. D.: Advances in the estimation of ice particle fall speeds using laboratory and field measurements, Journal of the Atmospheric Sciences, 67, 2469–2482, https://doi.org/10.1175/2010JAS3379.1, 2010.

Hinton, G. E. and Salakhutdinov, R. R.: Reducing the dimensionality of data with neural networks, Science, 313, https://doi.org/10.1126/science.1127647, 2006.

Hogan, R. J. and Westbrook, C. D.: Equation for the Microwave Backscatter Cross Section of Aggregate Snowflakes Using the Self-Similar Rayleigh–Gans Approximation, Journal of the Atmospheric Sciences, 71, 3292–3301, https://doi.org/10.1175/JAS-D-13-0347.1, 2014.

Hogan, R. J., Honeyager, R., Tyynelä, J., and Kneifel, S.: Calculating the millimetre-wave scattering phase function of snowflakes using the self-similar Rayleigh-Gans Approximation, Quarterly Journal of the Royal Meteorological Society, 143, 834–844, https://doi.org/10.1002/qj.2968, 2017.

Ioffe, S. and Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift Sergey, Proceedings of the 32nd International Conference on Machine Learning, 37, 448—-456, https://proceedings.mlr.press/v37/ioffe15.html, 2015.

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

Khain, A. P., Beheng, K. D., Heymsfield, A. J., Korolev, A., Krichak, S. O., Levin, Z., Pinsky, M., Phillips, V., Teller, A., van den Heever, S. C., and Yano, J.-I.: Reviews of Geophysics, Reviews of Geophysics, 53, 247–322, https://doi.org/10.1002/2014RG000468, 2015.

775 Kingma, D. P. and Ba, J. L.: Adam: A method for stochastic optimization, 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, pp. 1–15, 2015.

Kneifel, S., Kulie, M. S., and Bennartz, R.: A triple-frequency approach to retrieve microphysical snowfall parameters, Journal of Geophysical Research Atmospheres, 116, 1–15, https://doi.org/10.1029/2010JD015430, 2011.

Kneifel, S., Von Lerber, A., Tiira, J., Moisseev, D., Kollias, P., and Leinonen, J.: Observed relations between snowfall microphysics and 780 triple-frequency radar measurements, Journal of Geophysical Research, 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, Geophysical Research Letters, 43, 2225–2233, https://doi.org/10.1002/2015GL067618, 2016.

Kramer, M. A.: Nonlinear principal component analysis using autoassociative neural networks, AIChE Journal, 37, 785 https://doi.org/10.1002/aic.690370209, 1991.

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, Journal of Atmospheric and Oceanic Technology, 34, 2375–2392, https://doi.org/10.1175/JTECH-D-17-0019.1, 2017.

Kulie, M. S., Hiley, M. J., Bennartz, R., Kneifel, S., and Tanelli, S.: Triple-frequency radar reflectivity signatures of snow: Observations 790 and comparisons with theoretical ice particle scattering models, Journal of Applied Meteorology and Climatology, 53, 1080–1098, https://doi.org/10.1175/JAMC-D-13-066.1, 2014.

Kuo, K. S., Olson, W. S., Johnson, B. T., Grecu, M., Tian, L., Clune, T. L., Van Aartsen, B. H., Heymsfield, A. J., Liao, L., and Meneghini, R.: Full access the microwave radiative properties of falling snow derived from nonspherical ice particle models. Part I: An extensive database of simulated pristine crystals and aggregate particles, and their scattering properties, Journal of Applied Meteorology and Climatology, 795 55, 691–708, https://doi.org/10.1175/JAMC-D-15-0130.1, 2016.

Lawson, R. P. and Baker, B. A.: Improvement in determination of ice water content from two-dimensional particle imagery. Part II: Applications to collected data, Journal of Applied Meteorology and Climatology, 45, 1291–1303, https://doi.org/10.1175/JAM2399.1, 2006.

Leinonen, J., Kneifel, S., Moisseev, D., Tyynelä, J., Tanelli, S., and Nousiainen, T.: Evidence of nonspheroidal behavior in millimeter-wavelength radar observations of snowfall, Journal of Geophysical Research Atmospheres, 117, 1–10, 800 https://doi.org/10.1029/2012JD017680, 2012.

Leinonen, J., Kneifel, S., and Hogan, R. J.: Evaluation of the Rayleigh–Gans approximation for microwave scattering by rimed snowflakes, Quarterly Journal of the Royal Meteorological Society, 144, 77–88, https://doi.org/10.1002/qj.3093, 2018a.

Leinonen, J., Lebsock, M. D., Tanelli, S., Sy, O. O., Dolan, B., Chase, R. J., Finlon, J. A., Von Lerber, A., and Moisseev, D.: Retrieval of snowflake microphysical properties from multifrequency radar observations, Atmospheric Measurement Techniques, 11, 5471–5488, 805 https://doi.org/10.5194/amt-11-5471-2018, 2018b.

Leinonen, J., Grazioli, J., and Berne, A.: Reconstruction of the mass and geometry of snowfall particles from multi-angle snowflake camera (MASC) images, Atmospheric Measurement Techniques, 14, 6851–6866, https://doi.org/10.5194/amt-14-6851-2021, 2021.

Leroy, D., Fontaine, E., Schwarzenboeck, A., and Strapp, J. W.: Ice Crystal sizes in high ice water content clouds. Part I: On the computation of median mass diameter from in situ measurements, Journal of Atmospheric and Oceanic Technology, 33, 2461–2476, https://doi.org/10.1175/JTECH-D-15-0151.1, 2016.

Liao, L., Meneghini, R., Tokay, A., and Bliven, L. F.: Retrieval of snow properties for Ku- and Ka-band dual-frequency radar, Journal of Applied Meteorology and Climatology, 55, 1845–1858, https://doi.org/10.1175/JAMC-D-15-0355.1, 2016.

Liu, G.: Approximation of Single Scattering Properties of Ice and Snow Particles for High Microwave Frequencies, Journal of the Atmospheric Sciences, 61, 2441–2456, https://doi.org/10.1175/1520-0469(2004)061<2441:AOSSPO>2.0.CO;2, 2004.

Lu, Y., Jiang, Z., Aydin, K., Verlinde, J., Clothiaux, E., and Botta, G.: A polarimetric scattering database for non-spherical ice particles at microwave wavelengths, Atmospheric Measurement Techniques, 9, 5119–5134, https://doi.org/10.5194/amt-9-5119-2016, 2016.

Maahn, M., Turner, D. D., Löhnert, U., Posselt, D. J., Ebell, K., Mace, G. G., and Comstock, J. M.: Optimal estimation retrievals and their uncertainties, Bulletin of the American Meteorological Society, 101, E1512–E1523, https://doi.org/10.1175/BAMS-D-19-0027.1, 2020.

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, Journal of Geophysical Research: Atmospheres, 123, 13,807–13,834, https://doi.org/10.1029/2018JD028603, 2018.

Matrosov, S. Y.: A dual-wavelength radar method to measure snowfall rate, Journal of Applied Meteorology, 37, 1510–1521, https://doi.org/10.1175/1520-0450(1998)037<1510:ADWRMT>2.0.CO;2, 1998.

Matrosov, S. Y., Uttal, T., Snider, J. B., and Kropfli, R. A.: Estimation of Ice Cloud Parameters from Ground-Based Infrared Radiometer and Radar Measurements, Journal of Geo, 97, 567–574, 1992.

Matus, A. V. and L'Ecuyer, T. S.: The role of cloud phase in Earth's radiation budget, Journal of Geophysical Research, 122, 2559–2578, https://doi.org/10.1002/2016JD025951, 2017.

McFarquhar, G. M., Baumgardner, D., Bansemer, A., Abel, S. J., Crosier, J., French, J., Rosenberg, P., Korolev, A., Schwarzoenboeck, A., Leroy, D., Um, J., Wu, W., Heymsfield, A. J., Twohy, C., Detwiler, A., Field, P., Neumann, A., Cotton, R., Axisa, D., and Dong, J.: Processing of Ice Cloud In Situ Data Collected by Bulk Water, Scattering, and Imaging Probes: Fundamentals, Uncertainties, and Efforts toward Consistency, Meteorological Monographs, 58, 11.1–11.33, https://doi.org/10.1175/amsmonographs-d-16-0007.1, 2017.

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, Geoscientific Model Development, 13, 4229–4251, https://doi.org/10.5194/gmd-13-4229-2020, 2020.

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, vol. 12, https://doi.org/10.1029/2019MS001689, 2020.

Mroz, K., Battaglia, A., Nguyen, C., Heymsfield, A., Protat, A., and Wolde, M.: Triple-frequency radar retrieval of microphysical properties of snow, Atmospheric Measurement Techniques, 14, 7243–7254, https://doi.org/10.5194/amt-14-7243-2021, 2021.

Nguyen, C. M., Wolde, M., Battaglia, A., Nichman, L., Bliankinshtein, N., Haimov, S., Bala, K., and Schuettemeyer, D.: Coincident in situ and triple-frequency radar airborne observations in the Arctic, Atmospheric Measurement Techniques, 15, 775–795, https://doi.org/10.5194/amt-15-775-2022, 2022.

Noh, Y. J., Seaman, C. J., Vonder Haar, T. H., and Liu, G.: In situ aircraft measurements of the vertical distribution of liquid and ice water content in midlatitude mixed-phase clouds, Journal of Applied Meteorology and Climatology, 52, 269–279, https://doi.org/10.1175/JAMC-D-11-0202.1, 2013.

845 Ori, D., von Terzi, L., Karrer, M., and Kneifel, S.: snowScatt 1.0: consistent model of microphysical and scattering properties of rimed and unrimed snowflakes based on the self-similar Rayleigh–Gans approximation, Geoscientific Model Development, 14, 1511–1531, https://doi.org/10.5194/gmd-14-1511-2021, 2021.

PAMTRA team: PAMTRA, https://pamtra.readthedocs.io/en/latest/, 2015.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf,
850 A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S.: PyTorch: An imperative style, high-performance deep learning library, Advances in Neural Information Processing Systems, 32, 2019.

Plana-Fattori, A., Protat, A., and Delanoë, J.: Observing ice clouds with a Doppler cloud radar, Comptes Rendus Physique, 11, 96–103, https://doi.org/10.1016/j.crhy.2009.11.004, 2010.

Praz, C., Roulet, Y. A., and Berne, A.: Solid hydrometeor classification and riming degree estimation from pictures collected with a Multi-
855 Angle Snowflake Camera, Atmospheric Measurement Techniques, 10, 1335–1357, https://doi.org/10.5194/amt-10-1335-2017, 2017.

Rasmussen, R., Cole, J., Moore, R. K., and Kuperman, M.: Common snowfall conditions associated with aircraft takeoff accidents, Journal of Aircraft, 37, 110–116, https://doi.org/10.2514/2.2568, 2000.

Rees, K. N., Singh, D. K., Pardyjak, E. R., and Garrett, T. J.: Mass and density of individual frozen hydrometeors, Atmospheric Chemistry and Physics, 21, 14 235–14 250, https://doi.org/10.5194/acp-21-14235-2021, 2021.

860 Rodgers, Clive D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, World Scientific Publishing Co.Pte.Ltd., 2, 2000.

Schneebeli, M., Dawes, N., Lehning, M., and Berne, A.: High-resolution vertical profiles of X-band polarimetric radar observables during snowfall in the Swiss Alps, Journal of Applied Meteorology and Climatology, 52, 378–394, https://doi.org/10.1175/JAMC-D-12-015.1, 2013.

Schwarzenboeck, A., Heintzenberg, J., and Mertes, S.: Incorporation of aerosol particles between 25 and 850 nm into cloud ele-
865 ments: Measurements with a new complementary sampling system, Atmospheric Research, 52, 241–260, https://doi.org/10.1016/S0169-8095(99)00034-4, 2000.

Sharman, R. D., Cornman, L. B., Meymaris, G., Pearson, J., and Farrar, T.: Description and derived climatologies of automated in situ eddy-dissipation-rate reports of atmospheric turbulence, Journal of Applied Meteorology and Climatology, 53, 1416–1432, https://doi.org/10.1175/JAMC-D-13-0329.1, 2014.

870 Shupe, M. D., Kollias, P., Matrosov, S. Y., and Schneider, T. L.: Deriving Mixed-Phase Cloud Properties from Doppler Radar Spectra, Journal of Atmospheric and Oceanic Technology, 21, 660–670, https://doi.org/10.1175/1520-0426(2004)021<0660:DMCPFD>2.0.CO;2, 2004.

Shupe, M. D., Brooks, I. M., and Canut, G.: Evaluation of turbulent dissipation rate retrievals from Doppler Cloud Radar, Atmospheric Measurement Techniques, 5, 1375–1385, https://doi.org/10.5194/amt-5-1375-2012, 2012.

Straka, J. M.: Cloud and precipitation microphysics: principles and parameterizations, Cambridge University Press,
875 https://doi.org/https://doi.org/10.1017/CBO9780511581168, 2009.

Szyrmer, W. and Zawadzki, I.: Snow studies. Part IV: Ensemble retrieval of snow microphysics from dual-wavelength vertically pointing radars, Journal of the Atmospheric Sciences, 71, 1171–1186, https://doi.org/10.1175/JAS-D-12-0286.1, 2014.

Taszarek, M., Kendzierski, S., and Pilguj, N.: Hazardous weather affecting European airports: Climatological estimates of situations with limited visibility, thunderstorm, low-level wind shear and snowfall from ERA5, Weather and Climate Extremes, 28, 100 243,
880 https://doi.org/10.1016/j.wace.2020.100243, 2020.

Tetoni, E., Ewald, F., Hagen, M., Köcher, G., Zinner, T., and Groß, S.: Retrievals of ice microphysics using dual-wavelength polarimetric radar observations during stratiform precipitation events, Atmospheric Measurement Techniques Discussions, https://doi.org/10.5194/amt-2021-216, 2021.

Viltard, N., Le Gac, C., Martini, A., Lemaître, Y., Pauwels, N., Delanoë, J., and Lesage, G.: Développements radar au LATMOS pour l'études des propriétés microphysiques des nuages et des précipitations, Tech. rep., https://hal.archives-ouvertes.fr/hal-02389653, 2019.

Vogl, T., Maahn, M., Kneifel, S., Schimmel, W., Moisseev, D., and Kalesse-Los, H.: Using artificial neural networks to predict riming from Doppler cloud radar observations, Atmospheric Measurement Techniques, 15, 365–381, https://doi.org/10.5194/amt-15-365-2022, 2022.

Zeiler, M. D., Krishnan, D., Taylor, G. W., and Fergus, R.: Deconvolutional networks, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2528–2535, https://doi.org/10.1109/CVPR.2010.5539957, 2010.

885