The Cryosphere

Identification of blowing snow particles in images from a Multi-Angle Snowflake Camera

Mathieu Schaer, Christophe Praz, and Alexis Berne

Environmental Remote Sensing Laboratory, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Correspondence: Alexis Berne (alexis.berne@epfl.ch)

Abstract. A new method to automatically discriminate between hydrometeors and blowing snow particles on Multi-Angle Snowflake Camera (MASC) images is introduced. The method uses four selected descriptors related to the image frequency, the number of particles detected per image, and their size and geometry to classify each individual image. The classification task is achieved with a two-component Gaussian mixture model fitted on a subset of representative images of each class from field campaigns in Antarctica and Davos, Switzerland. The performance is evaluated by labeling the subset of images on which the model was fitted. An overall accuracy and a Cohen kappa score of 99.4 % and 98.8 %, respectively, are achieved. In a second step, the probabilistic information is used to flag images composed of a mix of blowing snow particles and hydrometeors, which turns out to occur frequently. The percentage of images belonging to each class from an entire austral summer in Antarctica and during a winter in Davos, respectively, is presented. The capability to distinguish precipitation, blowing snow and a mix of those in MASC images is highly relevant to disentangle the complex interactions between wind, snowflakes and snowpack close to the surface.

1 Introduction

Over snow-covered regions, ice particles can be lifted from the surface by the wind and suspended in the atmosphere. Wind-driven snow transport is ubiquitous in the cryosphere: over complex terrain (e.g., Winstral et al., 2002; Mott and Lehning, 2010), over tundra/prairies (e.g., Pomeroy and Li, 2000) and over polar ice sheets (e.g., Bintanja, 2001; Déry and Yau, 2002; Palm et al., 2011). Wind-driven snow transport must be taken into account to obtain accurate estimates of the mass balance and radiative forcings at the surface (e.g., Gallée et al., 2001; Lesins et al., 2009; Scarchilli et al., 2010; Yang et al., 2014). In mountainous regions, wind-transported snow also creates local accumulations and irregular deposits, being a critical factor influencing avalanche formation (e.g., Schweizer et al., 2003). Quantifying snow transport during snowfall events and subsequent periods of strong winds is essential for local avalanche prediction (e.g., Lehning and Fierz, 2008). In the context of climate change, the mass balance of the Antarctic ice sheet is of increasing relevance due to its impact on sea level rise (Shepherd et al., 2012). The sustained katabatic winds in Antarctica generate frequent blowing snow events, which remove a significant amount of new snow through transport and sublimation. Wind-transported snow is hence an important factor to take into account when considering Antarctic mass balance (e.g., Déry and Yau, 2002; Scarchilli et al., 2010; Lenaerts and van den Broeke, 2012; Das et al., 2013). Blowing snow is also an important process for the mass balance of the Greenland ice sheet (e.g., Box et al., 2006).

Ice particles moving at the snow surface belong to one of the three main types of associated motion: creep, saltation and suspension (e.g., Kind, 1990). Given the fact that the observations used in the present study were collected about 3 m above the ground (or snow surface) level, the term "blowing snow" hereinafter refers to wind-suspended ice particles.

Blowing snow is challenging to measure and characterize. Various approaches have been proposed to monitor blowing snow at ground level: mechanical traps, nets, photoelectric or acoustic sensors, and photographic systems (Leonard et al., 2012; Kinar and Pomeroy, 2015). Although not specifically designed for blowing snow, present weather sensors have

been shown to be valuable to monitor drifting and blowing snow fluxes (e.g., Bellot et al., 2011). Remote sensing and lidar systems in particular have recently been used to characterize the occurrence and depth of blowing snow layers, either from space (Palm et al., 2011) or near ground level (Gossart et al., 2017). Suspended ice particles are under the influence of the gravitational force proportional to the size cubed while the drag force is proportional to the area (size squared). With a greater area-to-mass ratio, smaller particles are thus more likely to be lifted in the suspension layer. A comparison of 10 different studies of measured and simulated particle size distributions of blowing snow reveals mean diameters at heights above 0.2 m ranging from 50 to 160 μm (Gordon and Taylor, 2009).

Blowing snow may also contaminate precipitation observations collected by ground-based sensors, frequently in Antarctica (e.g., Nishimura and Nemoto, 2005; Gossart et al., 2017), where winds are strong and frequent, but also in snowy regions in general (Rasmussen et al., 2012; Naaim-Bouvet et al., 2014; Scaff et al., 2015). The issue of snowfall measurement is complex, and the WMO promoted intercomparison projects to evaluate various sensors and define standard setups and protocols over the last two decades, as illustrated in Goodison et al. (1998) and the recent SPICE project (http://www.wmo.int/pages/prog/www/IMOP/intercomparisons/SPICE/SPICE.html, last access: 28 January 2020).

The Multi-Angle Snowflake Camera (MASC) is a ground-based instrument designed to automatically capture high-resolution ($\sim$ 33.5 μm) photographs of falling hydrometeors from three different angles (Garrett et al., 2012). The MASC has been used in previous studies to investigate snowflake properties (Garrett et al., 2015; Grazioli et al., 2017) and to help interpret weather radar measurements (Kennedy et al., 2018). Interestingly, blowing snow particles also trigger the MASC motion detector system (see Sect. 2.1), producing many images in windy environments. In addition to hydrometeor classification techniques based on MASC images (e.g., Praz et al., 2017), the ability to discriminate between images composed of blowing snow and precipitation particles would therefore be relevant to characterize blowing snow, to provide reference observations to improve its remote sensing and to obtain more accurate snowfall estimates from ground-based sensors. More generally, detailed information about the type of particles extracted from pictures collected by a MASC will enable us to further investigate the complex interactions between wind, snowflakes and snowpack close to the surface in cold and windy regions.

This article presents a new method to automatically determine if an image from the MASC (and potentially other imaging instruments) is composed of blowing snow particles, precipitating hydrometeors (snowflakes and ice crystals) or a mix of both. The classification is accomplished by means of a Gaussian mixture model (GMM) with two components, fitted on a set of representative MASC images and evaluated on a manually built validation set. The paper is organized as follows: Sect. 2 introduces the data sets used to develop the method and fit the GMM. Section 3 illustrates the different steps to isolate the particles and extract related features for the clustering task. Section 4 explains the selection of the most relevant features, the fitting of the GMM and the attribution of a flag for mixed images. The main results are shown in Sect. 5. At last, limitations and further improvements are discussed in Sect. 6.

2 Instrument and data sets

2.1 The Multi-Angle Snowflake Camera

The MASC is a ground-based instrument which automatically takes high-resolution and stereoscopic photographs of hydrometeors in free fall while measuring their fall velocity. Its working mechanism is only summarized hereafter, as more details and explanations can be found in Garrett et al. (2012), who provide an extensive description of the instrument. Three high-resolution cameras (2448 pixels $\times$ 2048 pixels), separated by an angle of 36°, are attached to a ring structure and form altogether the imaging unit (see Fig. 1). The focal point is located inside the ring at about 10 cm from each camera (with a focal length of 12.5 mm). Particles falling through the ring and detected by the two horizontally aligned near-infrared emitter–receiver arrays trigger the three flashes and the three cameras. The cameras' apertures and exposure times were adjusted in order to maximize the contrast on hydrometeor photographs while preventing motion blur effects, leading to a resolution of about 33.5 μm and a sampling area of about 8.3 cm^2 (see Praz et al., 2017). The maximum frequency of triggering is 3 Hz, which is three image triplets per second (see Fig. 6).

These specifications can be compared to the snow particle counter (SPC) which has been used in many studies of blowing snow (e.g., Nishimura and Nemoto, 2005; Gordon and Taylor, 2009; Kinar and Pomeroy, 2015; Guyomarc'h et al., 2019) and can be considered as the reference instrument for monitoring blowing snow (e.g., Crivelli et al., 2016). The SPC has a control volume of 2 mm $\times$ 25 mm $\times$ 0.5 mm and assigns particles into 32 diameter classes between 50 and 500 μm. It provides information on particle diameter (assuming a spherical shape), particle number and particle mass flux usually at a 1 s resolution (but raw data are measured at up to 150 kHz, Nishimura et al., 2014). For more information about the SPC, the reader is referred to the articles mentioned above.

2.2 Data sets

The MASC data used to implement and validate the present algorithm were collected during three field campaigns. The first one took place in Davos, Switzerland, from October 2015 to June 2016. The MASC was placed at 2540 m a.s.l

Figure 1. (a) Side view of the MASC with the three flash lamps in white on top and the two detectors as white boxes on the side of the metal ring (in black and red in front). **(b)** Top view of the inside of the MASC, with the three cameras clearly visible.

in a double-fence intercomparison reference (DFIR; see Fig. 2a), designed to limit the adverse effect of wind on the measuring instruments in its center (Goodison et al., 1998). The MASC was about 3 m above ground. The two other campaigns took place at the French Antarctic Dumont d'Urville (DDU) station, on the coast of Adélie Land, from November 2015 to February 2015 and from January to July 2017, in the framework of the Antarctic Precipitation, Remote Sensing from Surface and Space project (http://apres3.osug.fr, last access: 28 January 2020) (Grazioli et al., 2017; Genthon et al., 2018). The instrument was deployed on a rooftop at about 3 m above ground (see Fig. 2b). A collocated weather station and a micro rain radar (MRR) were also installed. Nearly 3 million images were collected during these measurement campaigns altogether.

From this large number of data, subsets of pure precipitation and pure blowing snow images were manually selected and further analyzed to choose relevant descriptors and fit a two-component GMM. The task of selecting a sufficient number of representative images for both classes turned out to be more complicated than expected, in particular for the Antarctic data set in which mixed images are very frequent. Gossart et al. (2017) used ceilometer data collected at the Neumayer (coastal) and Princess Elizabeth (inland) stations in East Antarctica to investigate blowing snow, and they suggest that more than 90 % of blowing snow occurs during synoptic events, usually combined with precipitation. For the sake of generalization, a large number of representative events were selected across the three campaigns. The goal was to cover a wide range of hydrometeors types and a wide range of snowfall intensities for the precipitation subset. Similarly, a wide range of wind speeds and concentrations were considered to build the blowing snow subset. From the campaigns in Antarctica, pure blowing snow and hydrometeor events were highlighted by comparing time series of MASC image frequency, wind speed and MRR-derived rain rate, as illustrated in Fig. 3. It was noticed that during strong blowing snow events, the number of images captured by the MASC was much larger than during precipitation events (more than one image per second; see Fig. 6). Potential pure blowing snow events were selected when the MASC image frequency and wind speed were higher than their respective median estimated over the whole campaign (to select relatively high values), and no precipitation was detected during the preceding hour. Only events for which these criteria applied for over an hour consecutively were kept. To highlight pure precipitation, the principle was the same but the criteria were an image frequency and a wind speed lower than the median as well as a MRR precipitation rate greater than zero. The MRR has a certain detection limit, so it was noticed that events selected as blowing snow could also occur during undetected light precipitation. As a result, images from all events were rapidly checked visually, and the campaign logbook was consulted to ensure that the selection was consistent and coherent. In both cases, some events had to be removed because of obvious mixing of blowing snow and hydrometeors.

As the MASC was deployed inside a DFIR in Davos, no blowing snow events were selected from this campaign. Although the DFIR is supposed to shelter the inner instruments from wind disturbances, we noticed that many images do not solely contain pure hydrometeors. From a webcam monitoring the instrumental setup, we noticed that the fresh snow accumulated on the edges and borders of the wooden structure of the DFIR was frequently blown away towards the sensor. To enlarge the precipitation subset, events with high snowfall rate but not affected by outliers of fresh wind-blown snow were added. Finally, some sparse images of obvious pure hydrometeors in the middle of mixed events were also included in the training set. In total, each subset contained 4263 images and, despite possible remaining (limited) uncertainty in the exact type of images, is assumed to be accurate and reliable enough to serve as reference for the evaluation of the proposed technique (see Fig. 8 and Sect. 4.2).
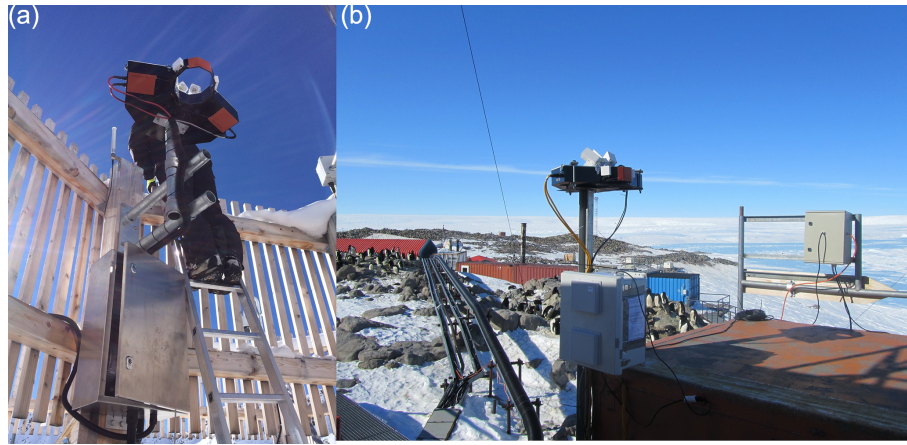
Figure 2. Experimental setup conditions of the MASC in a DFIR near Davos **(a)** and on top of a container at Dumont d'Urville **(b)**.
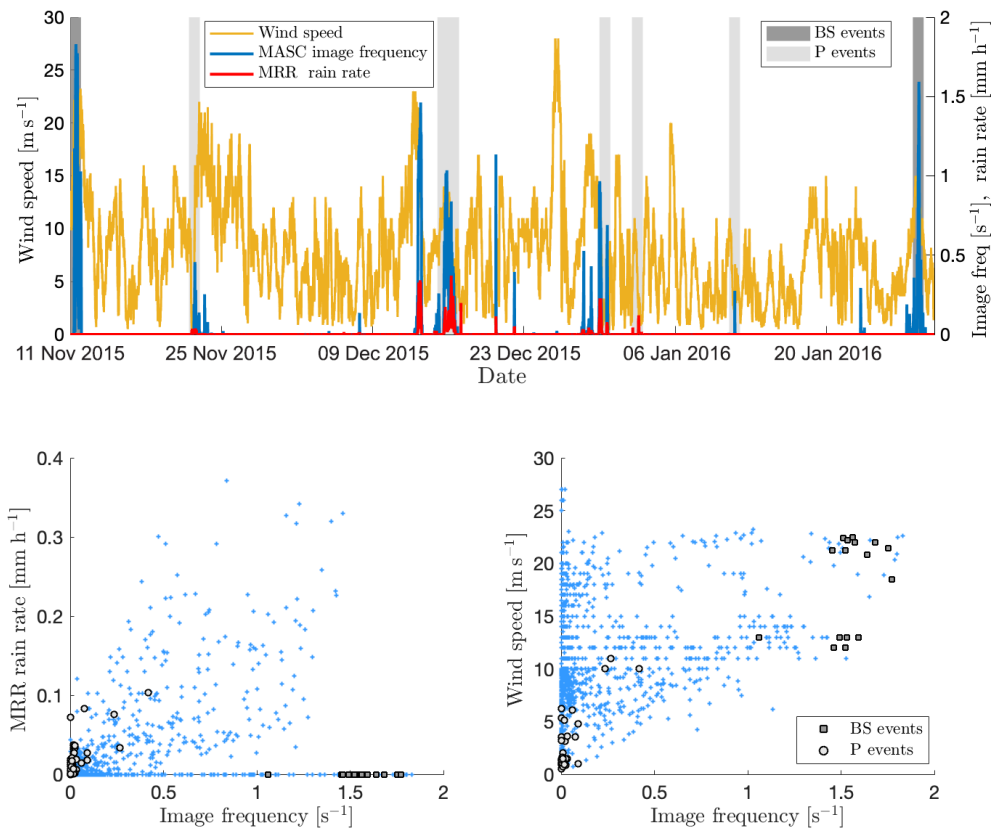

Figure 3. Time series and scatter plots of MASC image frequency, wind speed (measured at 10 m) and MRR-derived rain rate for the Antarctica 2015–2016 campaign. The gray shading indicates days during which time steps have been selected for the training set as blowing snow (dark gray) or precipitation (light gray). In the bottom scatter plots, the markers figure the selected blowing snow and precipitation time steps. Points on the x axis in the left scatter plot are potential candidates for pure blowing snow.

3 Image processing

3.1 Particle detection

The MASC instrument and the collected images are described in Sect. 2.1. Although a single particle activates the cameras, many MASC pictures contain multiple particles distributed over the entire image, especially when blowing snow occurs. In fact, the number of particles appearing on a single image is a key characteristic to distinguish between precipitation and blowing snow. As a result, it was deemed

Table 1. Campaigns and dates of selected events for the blowing snow (BS) and precipitation (P) subsets.

Antarctica 2015–2016	Antarctica 2017	Davos 2015–2016
11 Nov BS	8 Feb BS	23 Feb P
22 Nov P	9 Feb BS	25 Feb P
15 Dec P	18 Feb BS	4 Mar P
16 Dec P	19 Feb BS	5 Mar P
30 Dec P		16 Mar P
2 Jan P		25 Mar P
11 Jan P		
28 Jan BS		

essential to detect all particles in each image rather than the triggering one only (which is sometimes unidentifiable). A key challenge of this approach was to get rid of the noisy background. For this purpose, a median filter was used. The brightness of the background strongly depends on the luminosity at the instant of the picture, which varies according to the time of day and can change abruptly in partly cloudy conditions when the sun suddenly appears from behind a cloud. As a result, the median filter shows better performance to remove the background when systematically recomputed over a small number of consecutive images. Assuming that snow particles rarely appear at the exact same position in several consecutive images, the median filter was chosen to be computed over blocks of five images per camera angle. To ensure complete removal of the background when its brightness is greater that the corresponding median, a factor of 1.1 was applied to the filter. Finally, as some limited residual noise can still remain in the filtered image, a small detection threshold of 0.02 grayscale intensity was applied to isolate the snow particles. Masks of the sky and reflecting parts of the background (i.e., metallic plates) were created for each camera. The multiplication factor and detection threshold are increased in the regions delineated by the masks if the normal filtering leads locally to more pixels detected that one can expect from real particles. These steps are illustrated in Figs. 4 and 5. Issues in the filtering may occur if consecutive images are separated by a period of time during which the ambient luminosity has changed significantly (e.g., before/after the sunrise or sunset). An example is shown in Fig. B1 in Appendix B.

3.2 Feature extraction

Machine-learning algorithms require a set of variables, commonly called features or descriptors, upon which the classification is performed. Because of the fragmentation of ice crystals when hitting the snow surface (e.g., Schmidt, 1980; Comola et al., 2017), blowing snow is expected to be characterized by much smaller particle size and much higher particle concentration than snowfall (e.g., Budd, 1966; Budd et al., 1966; Nishimura and Nemoto, 2005; Naaim-Bouvet
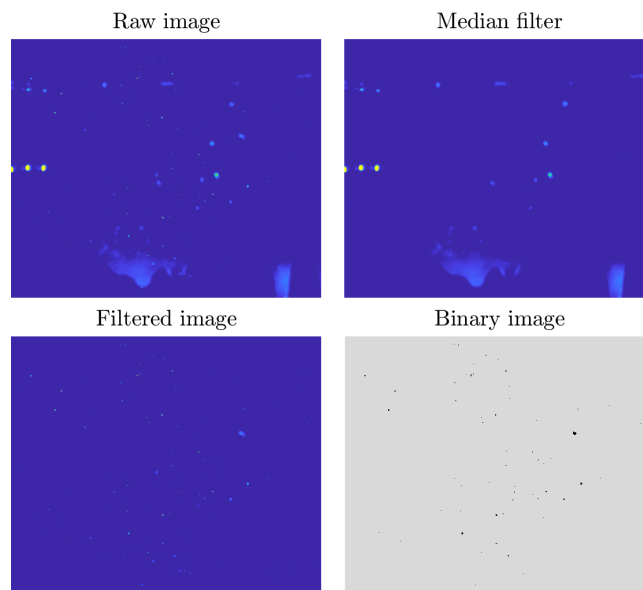
Figure 4. Raw image, median filter, filtered image and final binary image for an example of blowing snow particles. The image size is 2448 pixels × 2048 pixels, corresponding to 82 mm × 68.6 mm. Original MASC images are in gray shades, but the color scheme used here aims to enhance contrast and details for visual purposes.
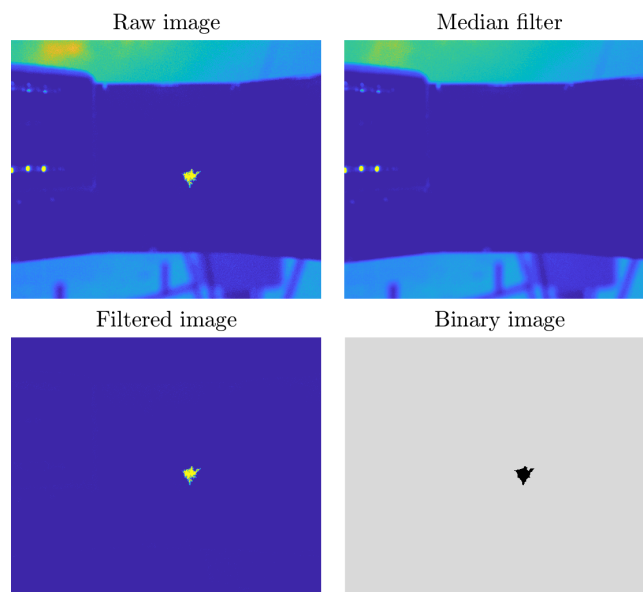
Figure 5. Raw image, median filter, filtered image and final binary image for an example of a hydrometeor. The color scheme is used to enhance details for visual purposes.

et al., 2014). In this study, various quantitative descriptors were therefore calculated according to four different categories: the number of particles and their spread across the image, the size of the particles, the geometry of the particles, and the frequency at which the images are taken.

Since it is difficult to exactly guess which descriptors are the most adequate to differentiate between blowing snow and precipitation images, an extensive collection of features was extracted from the blowing snow and precipitation subsets and compared. The selection of the most relevant ones is explained in the next section. As the classification is performed at the image level, we need features at the same level, and the information on the geometry and size of each detected particle in the considered image must hence be transformed into a single descriptor for that image. Consequently, quantiles ranging from 0 to 1 and moments from 1 to 10 were computed out of the distribution of the considered feature within the image. The image frequency is a descriptor independent of the content of the image and thus from the detection of particles. It is therefore not affected by potential image-processing issues. As each image comes with its attributed timestamp, the average number of images per minute was calculated with a moving window. The full list of all computed descriptors is displayed in Appendix A. The extraction of features was conducted with the MATLAB Image Processing Toolbox, in particular the function `regionprops` (https://ch.mathworks.com/help/images/ref/regionprops.html, last access: 28 January 2020).

4 Classification

4.1 Feature selection and transformation

Selecting a relevant set of features and avoiding redundancy is essential for accurate classification, regardless of the classification algorithm. For each of the four categories of descriptors previously mentioned, the most relevant one (according to the criterion explained below) was kept. The descriptor maximizing the interclusters-over-intraclusters distance described in Eq. (1) was selected. This quantity represents the distance between the mean of the blowing snow and precipitation distributions (μ_{BS} and μ_{P} respectively), normalized by the sum of their respective standard deviations (σ_{BS} and σ_{P} respectively).

$$S = \frac{|\mu_{\mathrm{BS}} - \mu_{\mathrm{P}}|}{\frac{1}{2}(\sigma_{\mathrm{BS}} + \sigma_{\mathrm{P}})}. \tag{1}$$

For the features describing the number of detected particles and their spread across the image, the cumulative distance transform was kept. It represents the sum over each entry of the distance transform matrix (https://ch.mathworks.com/help/images/ref/bwdist.html, last access: 28 January 2020) of the binary image. The distance transform matrix has the same dimensions as the binary image and computes, for each pixel, the Euclidean distance to the nearest 1 element (i.e., the nearest particle). As a result, an image with many particles well distributed over its entire surface will have a low cumulative distance transform, while a single

Table 2. Selected features and corresponding S values.

Feature name	S
Image frequency	4.43
Cumulative distance transform	2.89
Maximum diameter quantile 0.7	1.71
Squared fractal index quantile 0	3.81

gle particle, even particularly large, will have a high value. This descriptor is more robust to image-processing issues than the raw number of particles, as illustrated in Fig. B2 in Appendix B.

Concerning the size distribution of the particles detected in an image, the quantile 0.7 of the maximum diameter was selected (because it has the highest S value among the different quantiles tested). The maximum diameter (D_{max}) represents the longest segment between two edges of a particle (see Praz et al., 2017, for more details). A logarithmic transformation of this feature was performed to make the distributions of the two classes more Gaussian. The minimum (i.e., quantile 0) squared fractal index showed the greatest S value (hence discrimination potential) among the features related to the particle geometry indices. The fractal index (FRAC) is defined according to the formula proposed by McGarigal and Marks (1995) in the context of landscape-pattern analysis. It was also more recently used to quantify stand structural complexity from terrestrial laser scans of forests (Ehbrecht et al., 2017).

Due to its different nature, the image frequency descriptor was selected by default, but it is worth noting that it has the highest S value (Eq. 1) among all descriptors (Table 2). The marginal distributions of the selected descriptors for the training set are shown in Fig. 6 to provide an idea of their respective magnitude and variability, as well as to illustrate their discrimination potential. As noted above, the image frequency is the most informative descriptor to distinguish blowing snow and precipitation.

In summary, four descriptor categories (related to particle size, particle geometry and particle distribution within the image, and image frequency) have been defined to distinguish images collected during blowing snow or snowfall, based on the expected differences in particle size and concentration between the two. A number of descriptors were estimated from each image by computing various quantiles and moments of the distributions of geometric properties of the particles in the considered image. One descriptor from each of the four categories defined above (listed in Table 2) was then selected to be further used for classification as the one maximizing the interclusters-over-intraclusters distance defined in Eq. (1).
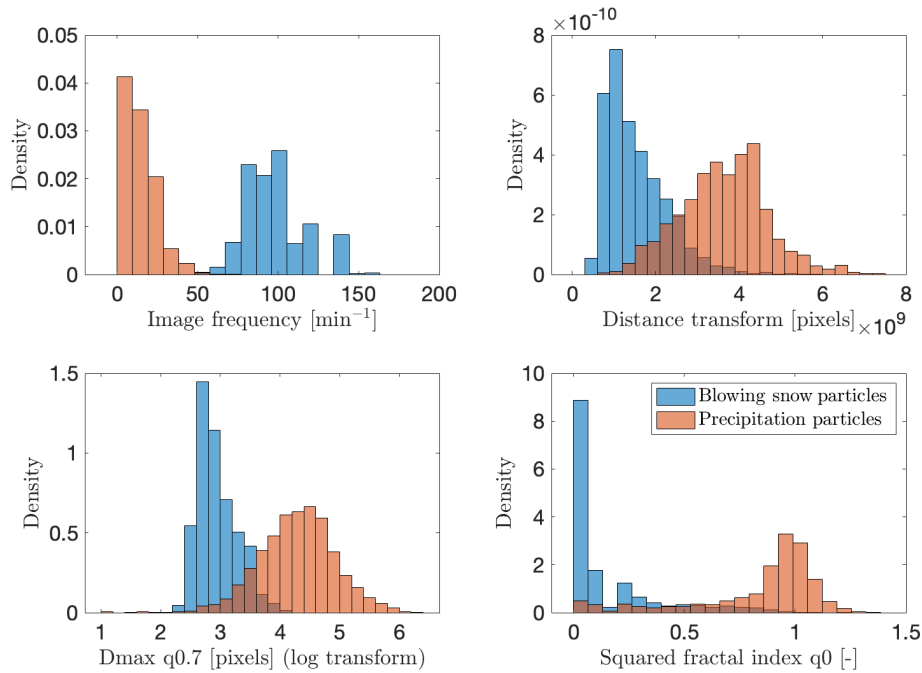
Figure 6. Histograms of selected descriptors for the training blowing snow and precipitation subsets.

4.2 Model fitting

The choice for the binary classification task was made on a Gaussian mixture model, an unsupervised learning technique that fits a mixture of multivariate Gaussian distributions to the data (see Murphy, 2012; McLachlan and Basford, 1988; Moerland, 2000, for more details). The mathematical description of a multivariate normal distribution is provided in Eq. (2).

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}}$$
$$\exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}, \quad (2)$$

where $\boldsymbol{x}$ is a Gaussian multivariate random variable of dimension D, $\boldsymbol{\mu}$ its mean and $\boldsymbol{\Sigma}$ its covariance matrix, with ^{T} the transpose operator.

The choice of an unsupervised approach is based on several reasons. First, unsupervised methods do not depend upon labels. Hence, it is not required to ensure correct labeling of each image in the training set. As mentioned earlier, many images are composed of a mix of blowing snow and precipitation, and it is thus difficult to guarantee the objectivity of all given labels. Second, a clear separation observed between the two subsets would be statistically highly significant as no prior information is provided to the learning algorithm about the classes. Third, for low-dimension problems, unsupervised methods are sometimes less prone to overfitting and have a better potential for generalization. A main advantage of the GMM compared to other unsupervised methods is to

provide posterior probabilities on the cluster assignments and thus allow for soft clustering (i.e., probabilistic assignment). In the context of the present study, this is absolutely relevant as there exists a whole continuum of in-between cases of mixed images. It should be noted that the descriptors were selected using a reference set (see previous section), but the clustering conducted by means of the GMM is itself unsupervised.

A two-component GMM with unshared full covariance matrices was thus fitted to the four-dimensional ($\boldsymbol{x} = \{f_1, f_2, f_3, f_4\}$, where f_i represents the four features listed in Table 2) data composed of the blowing snow and precipitation subsets. The MATLAB Statistics and Machine Learning Toolbox was used for this purpose, and the model parameters were estimated by maximum likelihood via the expectation–maximization (EM) algorithm (https://ch.mathworks.com/help/stats/gaussian-mixture-models-2.html, last access: 28 January 2020). The features were standardized before fitting the model. The mixing weights (or component proportions) were artificially set to 0.5 by randomly removing 80 data points from the training set and fitting again the GMM to have perfectly balanced classes. This step is essential as the model will then be used to classify new images (possibly from other campaigns). There are no reasons to give more weight to one component, as the relative proportion of blowing snow and precipitation images strongly depends on the campaign location. The posterior probabilities are computed using the Bayes rule (Murphy,

2012):

$$P(z_i = k | \boldsymbol{x}_i, \boldsymbol{\theta}) = \frac{P(\boldsymbol{x}_i | z_i = k, \boldsymbol{\theta}) P(z_i = k | \boldsymbol{\theta})}{P(\boldsymbol{x}_i | \boldsymbol{\theta})}, \qquad (3)$$

where z_i is a discrete latent variable taking the values $1, \ldots, K$ and labeling the K Gaussian components. $P(z_i = k | \boldsymbol{x}_i, \boldsymbol{\theta})$ is the posterior probability that point i belongs to cluster k (also known as the "responsibility" of cluster k for point i). $P(\boldsymbol{x}_i | z_i = k, \boldsymbol{\theta})$ corresponds to the density of component k at point i (i.e., $\mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$), and $P(z_i = k | \boldsymbol{\theta})$ represents the mixing weight (also denoted π_k). Note that the π_k's are positive and sum to 1. $\boldsymbol{\theta}$ refers to the fitted parameters of the mixture model $\{ \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K, \pi_1, \ldots, \pi_K \}$. $P(\boldsymbol{x}_i | \boldsymbol{\theta})$ is the marginal probability at point i, which is simply the weighted sum of all component densities:

$$P(\boldsymbol{x}_i | \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \qquad (4)$$

As the concern of this study is on two components only, a more compact notation will be used for the rest of the article. The latent variable z will be replaced by k_P and k_{BS} to refer to the precipitation and blowing snow clusters, respectively. The term $\boldsymbol{\theta}$, which denotes the model parameters, will be left implicit. Assuming we are at first interested in performing some hard clustering (i.e., single label to a given image), an image will be classified as blowing snow if $P(k_{BS} | \boldsymbol{x}_i) > P(k_P | \boldsymbol{x}_i)$. That is to say, if the posterior probability of belonging to the blowing snow cluster is greater than 0.5, an image will be classified as such (because the posterior probabilities sum to 1). The model performance was assessed by simply labeling the data points according to its initial subset. An overall accuracy of 99.4 % and a Cohen kappa score of 98.8 % were achieved. The Cohen kappa statistic adjusts the accuracy by accounting for correct predictions occurring by chance (Byrt et al., 1993). These high values indicate a very good performance of the fitted GMM. Figure 7 presents the fitted Gaussian components as well as the reference values (not used in the fitting) for each of the six possible pairs of the four descriptors. It clearly illustrates the performance of the fitted GMM and the discriminative power of the descriptor related to image frequency.

To investigate the stability of the Gaussian components, the precipitation and blowing snow subsets were both randomly permuted and divided into 10 equal parts. Ten new training sets of a balanced amount of each subset were created, and a new GMM was fitted. Figure 8 shows on the top line the boxplots of the Gaussian component parameters $\boldsymbol{\mu}_d$ and $\boldsymbol{\sigma}_d$ (i.e., diagonal entries of $\boldsymbol{\Sigma}$) for each of the four dimensions. The boxplots show a limited variability for each feature (below 10 %), indicating a reasonable stability of the fitted parameters. In addition, the bottom line of Fig. 8 presents the learning curves, and their fast convergence to the same horizontal line when more than 30 % of the training set

is used indicates a data set large enough for a reliable fitting of the GMM, without overfitting.

4.3   Flag for mixed images

As mentioned earlier, an asset of using a GMM model is the posterior probabilistic information that could help estimate the degree of mixing of an image. Data points located close to the decision boundary in the multidimensional space are likely to be composed of a mix of blowing snow particles and hydrometeors. However, distributions of posterior probabilities computed over thousands of new images from entire campaigns appeared to be stretched out on both ends of the domain (i.e., close to 0 or 1), and not many images were present in between. This is probably due to the nature of the descriptors and the resulting shapes and relative positions of the Gaussian distributions. In order to investigate this issue, an additional set of images corresponding to mixed cases was built: it exhibited clear differences in the posterior probabilities with the pure blowing snow and pure precipitation subsets. This differentiation was however around 10^{-6} (or $1 - 10^{-6}$), which is not so informative as such. Consequently, it was decided to define a new index similar to the posterior probability of belonging to the blowing snow component but more evenly distributed across the range $]0, 1[$. The new index uses the negative logarithm of the posterior probabilities multiplied by the marginal probability. Taking the log of Eq. (3) for k_{BS}, we have (the same applies for k_P)

$$-\log[P(k_{BS} | \boldsymbol{x}_i) P(\boldsymbol{x}_i)] = -\log[P(\boldsymbol{x}_i | k_{BS}) P(k_{BS})]. \qquad (5)$$

Noting that the term $P(\boldsymbol{x}_i | k_{BS})$ on the right-hand side is $\mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_{BS}, \boldsymbol{\Sigma}_{BS})$, one can substitute Eq. (2) into the above expression, which yields

$$-\log[P(k_{BS} | \boldsymbol{x}_i)] - \log[P(\boldsymbol{x}_i)] =$$
$$\frac{1}{2}(\boldsymbol{x}_i - \mu_{BS})^T \boldsymbol{\Sigma}_{BS}^{-1}(\boldsymbol{x}_i - \mu_{BS}) + \frac{1}{2}\log(|\boldsymbol{\Sigma}_{BS}|)$$
$$+ \frac{D}{2}\log(2\pi) - \log(P(k_{BS})). \qquad (6)$$

The quadratic term on the right-hand side is the Mahalanobis distance, which is a distance that uses a Σ^{-1} norm. Hence, it represents the distance between point $\boldsymbol{x}_i$ and the center of the distribution, corrected for correlations and unequal variances in the feature space (De Maesschalck et al., 2000). The second term is related to the determinant of the covariance matrix and equals -3.94 for the blowing snow component and -2.59 for the precipitation one. The two last terms are constant and sum to 4.37 (the component proportions were set to 0.5 and $D = 4$). The right side of Eq. (6) is also known as the quadratic discriminant function (QDF, Kimura et al., 1987), commonly noted $g_k(\boldsymbol{x}_i)$. The minus in front of the logarithm on the left side of Eq. (6) is used to return positive values and facilitate subsequent graphical interpretations. Note that the constant term $\frac{D}{2}\log(2\pi)$ is often
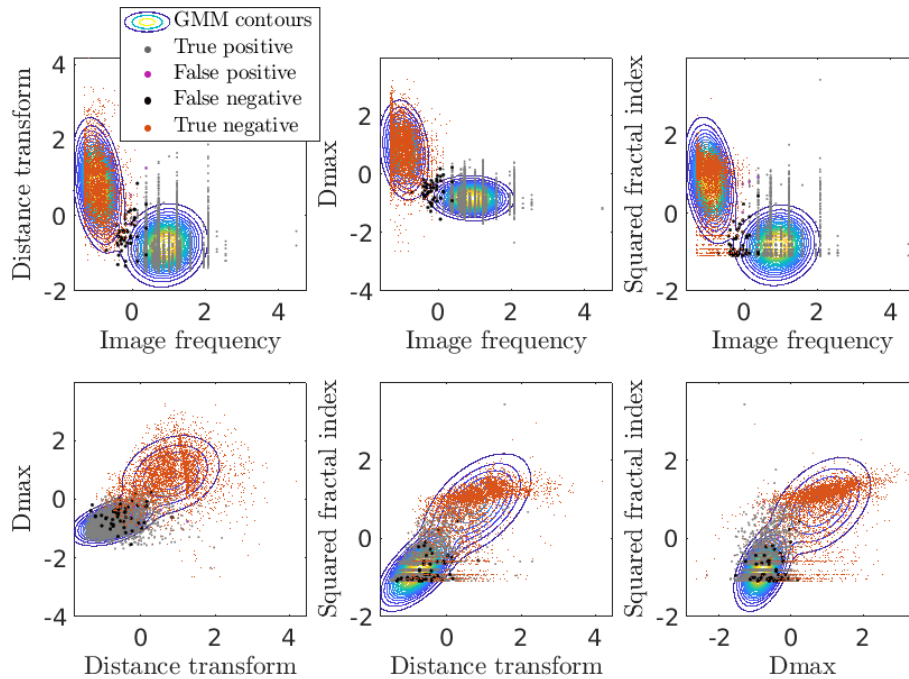
Figure 7. GMM contours and data points projected on the 2-D planes. The colors correspond to the four entries of the confusion matrix. The predictions result from the clustering and the ground truth is the given labels.

removed, but in this case it ensures that $g_k(\boldsymbol{x})$ is positive, even for a Mahalanobis distance of zero. Figure 9 displays a scatter plot of the quadratic discriminant values of both components for the whole training set. The proposed index is defined as the angle of the vector representing a data point on the scatter plot, normalized by $\frac{\pi}{2}$. It is thus computed as follows:

$$\psi = \frac{2}{\pi} \arctan \left\{ \frac{-\log[P(k_{\mathrm{P}}|\boldsymbol{x}_i)P(\boldsymbol{x}_i)]}{-\log[P(k_{\mathrm{BS}}|\boldsymbol{x}_i)P(\boldsymbol{x}_i)]} \right\}. \tag{7}$$

This normalized angle is bounded in $]0, 1[$, with values close to 1 (respectively 0) indicating a strong membership of the considered image (and not the respective proportions within this image) to the blowing snow (respectively precipitation) clusters. It is closely related to the asymmetry of the Mahalanobis distances between a point $\boldsymbol{x}_i$ and the centers of the two Gaussian distributions but corrected by the term $\frac{1}{2}\log(|\Sigma|)$, which is different for the two components. The advantage of using the index in this form, rather than deriving it from the Mahalanobis distances alone, is to respect the decision boundary given by the maximum a posteriori (MAP) rule. This means a posterior probability of 0.5 yields a ψ index of 0.5. Finally, quantiles 0.9 ($\psi_{\mathrm{P0.9}}$) and 0.1 ($\psi_{\mathrm{BS0.1}}$) of the ψ index distributions of the points classified as precipitation and blowing snow, respectively, are retained as thresholds to flag potential mixed images. The idea is to allow, for both classes, 10 % of the training set images to be flagged as mixed. This value is qualitatively supported by the distribution shown in Fig. 9. It can be changed by the user to be more

(increasing it) or less (decreasing it) strict on the classification as pure blowing snow or pure precipitation, depending on the intended application.

To ease reading and interpretation, a mixing index λ_{m} is introduced by linearly rescaling between 0 and 1 the ψ index of the images flagged as mixed (i.e., λ_{m} is not defined for pure precipitation or pure blowing snow images):

$$\begin{aligned} \lambda_{\mathrm{m}} \quad &= \tfrac{0.5}{0.5-\psi_{\mathrm{P0.9}}} \left(\psi - \psi_{\mathrm{P0.9}} \right) && \text{if } \psi \in (\psi_{\mathrm{P0.9}}, 0.5), \\ &= \tfrac{0.5}{0.5-\psi_{\mathrm{BS0.1}}} \left((1 - \psi_{\mathrm{BS0.1}}) - \psi \right) && \text{if } \psi \in [0.5, \psi_{\mathrm{BS0.1}}). \end{aligned} \tag{8}$$

The mixing index also respects the hard clustering assignment boundary at 0.5: $\lambda_{\mathrm{m}} > 0.5$ indicates that the image contains a mix of blowing snow and precipitation particles, but it is closer overall to blowing snow and vice versa. Images with a normalized angle outside the two mixed thresholds have an undefined (NaN) index of mixing and are considered as pure blowing snow particles or pure hydrometeors. Results are provided treating all images independently, but the ψ index can also be averaged among the three camera angles to provide a unique value per image identifier as well. The median of the range (max – min) covered by the ψ values from the three individual views is about 0.08 in Davos and 0.05 at Dumont d'Urville, indicating a limited variability between the three views.

In summary, the classification as a mixed case is based on the angle characterizing the considered MASC image in the 2-D space formed by the axis related to pure blowing snow on the one hand and the one related to pure precipitation on the other hand. A mixing index λ_{m} is finally computed by lin-
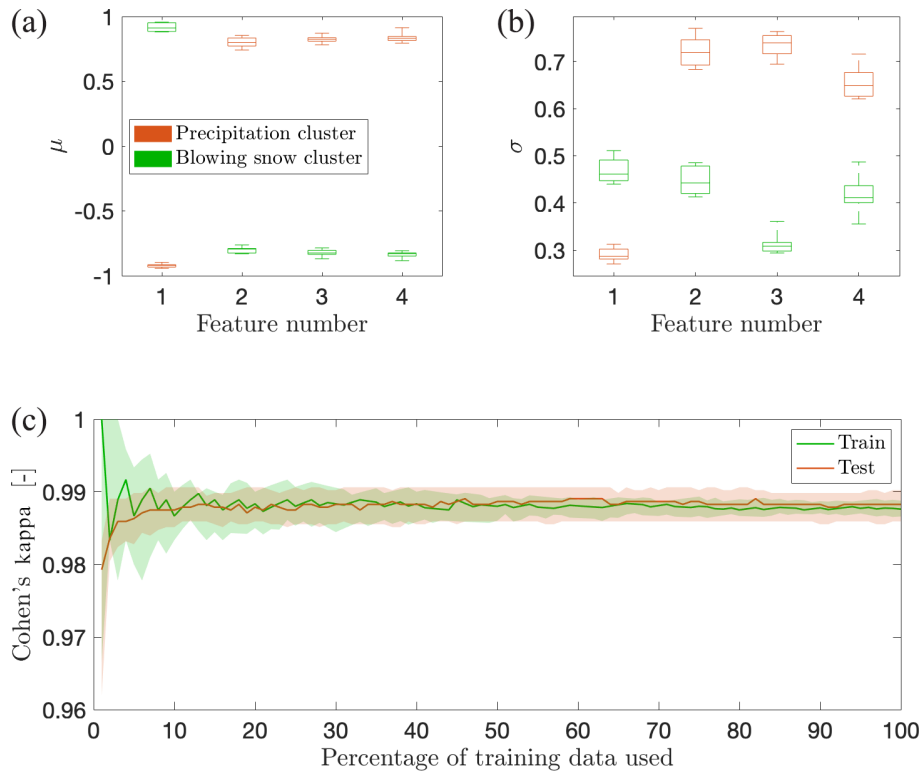
Figure 8. (a, b) Stability of the parameters μ and σ (diagonal entries of Σ) for the two Gaussian components. The boxplots show the distributions of these parameters for each dimension, after fitting the GMM on a 10-fold random split of the training set. The feature number follows the order given in Table 2. **(c)** Learning curves for the fitted GMM, showing the evolution of the train and Cohen's kappa test as a function of the proportion of the training samples used. The shaded areas correspond to the 25–75th percentile range computed over 40 iterations of 70 %–30 % random train-test splitting, and the bold lines are the medians.

early rescaling the normalized angle over the range of values corresponding to mixed cases.

5 Results

The method presented (and fitted) in the previous sections is now applied to the entire Antarctica 2017 campaign (January–July 2017) and to the entire Davos campaign (December 2015–March 2016). About 2×10^6 images for Antarctica and 8.5×10^5 for Davos were classified. Table 3 summarized the outcome in terms of respective proportions of pure blowing snow, pure precipitation, mixed blowing snow, and mixed precipitation for the Antarctic and Alpine data sets. As expected, the occurrence of blowing snow (pure + mixed) is much more frequent at Dumont d'Urville (75.6 %) than at Davos (21.5 %, out of which only 0.6 % if of pure blowing snow).

Figure 10a shows the distribution of the collected MASC images in the space formed by the two quadratic discriminant (one for blowing snow, one for precipitation), and Fig. 10b shows the distribution of the normalized angle for the entire Antarctica 2017 campaign. A clear difference with Fig. 9 is the large proportion of values corresponding to mixed cases:

Table 3. Percentages of MASC images per category.

Class	Antarctica (Jan–Jul 2017)	Davos (Dec 2015–Mar 2016)
Pure blowing snow	36.5 %	0.6 %
Pure precipitation	7.2 %	39.2 %
Mixed blowing snow	39.1 %	20.9 %
Mixed precipitation	17.2 %	39.3 %

there are much more points around the one–one line (Fig. 9a) and a small mode around 0.5 (Fig. 9b) for the entire campaign than for the training set (built with much fewer mixed cases). It is also clear from Fig. 10b that blowing snow and mixed blowing snow are more frequent than precipitation and mixed precipitation.

Figure 11 is similar to Fig. 10 but for the entire Davos data set. In comparison with Fig. 10, the occurrence of precipitation is much larger (and blowing snow much smaller), which is to be expected given the difference in geographic context (Alps vs. Antarctica) and experimental setup (wind-protected vs. no wind shield). It should be noted that mixed cases are relatively frequent and that blowing snow still hap-
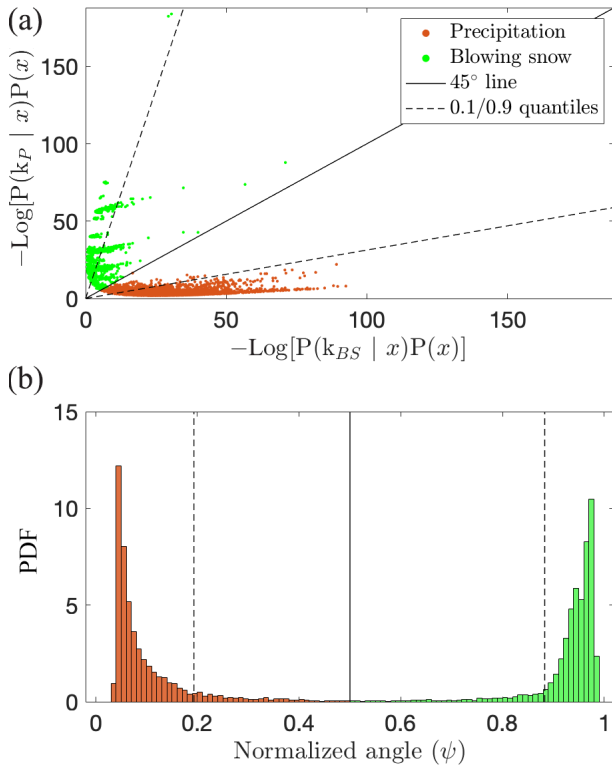
Figure 9. (a) Scatter plot of the quadratic discriminant values of both components for the training set. **(b)** Probability density functions (PDF) of the normalized angle for the precipitation and blowing snow subsets and thresholds to identify mixed images.

Figure 10. (a) Scatter plot of the quadratic discriminant values of both components for the entire Antarctica 2017 campaign. **(b)** Distribution of the normalized angle and corresponding classification.

pens in Davos although the MASC was located in a wind shielding fence.

Beyond global statistics on various data sets as presented above, the proposed approach can also be used to investigate the type of images at high temporal resolutions. Figure 12 shows an example of the output of the algorithm and corresponding images for a few time steps during a mixed event. It illustrates the capability of the proposed approach to distinguish blowing snow, precipitation and mixture in individual MASC images separated by a few seconds (and hence the contribution of the features other than image frequency). Over a longer time period, Fig. 13 displays the evolution of the normalized angle for a mixed event during the Antarctica 2017 campaign. From roughly 09:00 to 12:00, the types precipitation and mixed are dominant, while between 12:00 and 14:00 the three types (precipitation, mixed and blowing snow) occur simultaneously. This is to be expected at DDU where katabatic winds blow very frequently, even during precipitation (e.g., Vignon et al., 2019). From 14:00 to 22:00, blowing snow becomes dominant (because of stronger winds). After 22:00, mixed cases dominate and some images corresponding to precipitation are detected towards the end of the event. The possibility of identifying MASC images corresponding to precipitation, blowing snow or a mixture at
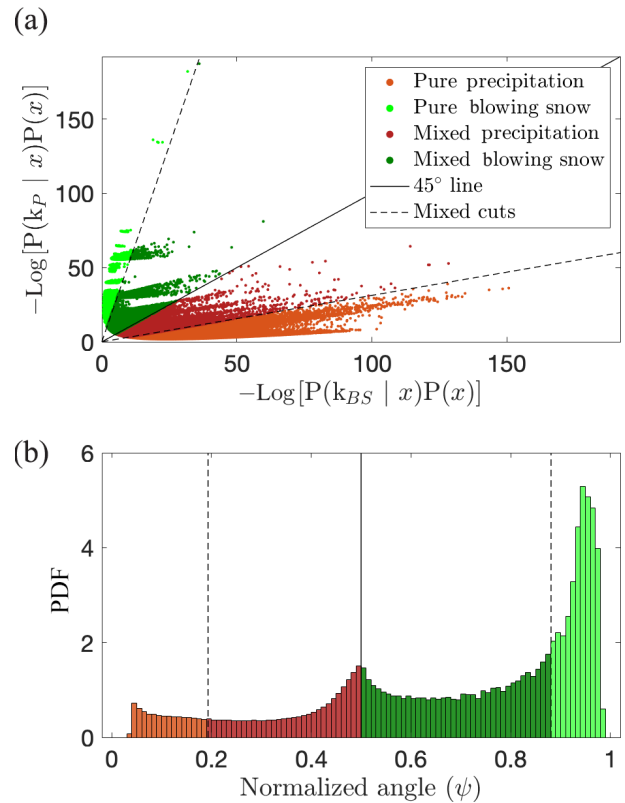
a temporal resolution high enough to capture the dynamics of the event is an interesting feature for regions where both are frequently associated.

Considering the full Antarctic and Alpine data sets, it is interesting to analyze the potential differences in their characteristics. Figure 14 presents the distributions of the four descriptors as in Fig. 6 but estimated from the entire data sets and not only the training sets (for images classified as pure blowing snow or pure precipitation). It can be seen that while the differences are limited for precipitation (slightly more frequent and larger in Davos than in Dumont d'Urville), they are significant for blowing snow: the blowing snow particles appear less fragmented (larger size and fractal index), less scattered within the images (larger distance transform) and with lower image frequencies in Davos. It should be recalled that the MASC was located in a wind-protecting fence in Davos, so first the occurrence of blowing snow is much smaller (0.6 vs. 36.5 %), and second it is likely related to fresh snow blown away from the top of the nearby fence.

The MASC resolution (33.5 µm) and thresholding (minimum 3 pixels in area) during image processing lead to an image resolution not high enough to capture in full detail the geometry of blowing snow particles. It is nevertheless interesting to plot the distribution of the measured sizes (asso-
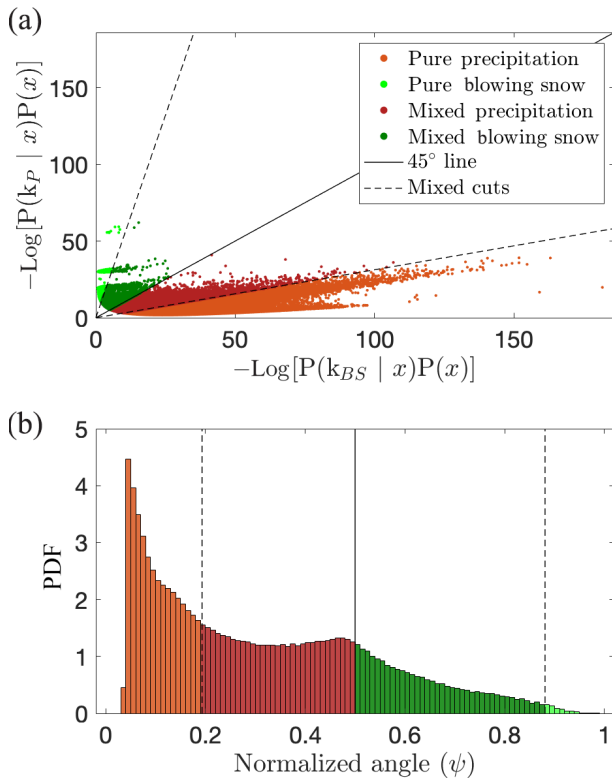
(a)

(b)

Figure 11. Same as Fig. 10 for the entire Davos campaign.

Timestamp	ID	Label	Normalized angle	Mixing index
25/03/2016 16:56:35	7383	1	0.7422	0.8178
25/03/2016 16:56:39	7384	0	0.1590	NaN
25/03/2016 16:56:39	7385	1	0.5659	0.5865
25/03/2016 16:56:40	7386	0	0.1437	NaN
25/03/2016 16:56:42	7387	0	0.1920	NaN
25/03/2016 16:56:42	7388	0	0.2260	0.0537

Figure 12. Consecutive MASC images from Davos and their respective classification label, normalized angle and mixing index. Label 1 is for blowing snow. A NaN mixing index means pure hydrometeor (or pure blowing snow). A mixing index close to 1 (top-left image) means that it is near pure blowing snow, while a value close to 0 (bottom-right image) indicates proximity to pure precipitation.

Figure 13. Time series of classified MASC images and corresponding ψ values (averaged over the three views) for a mixed event during the Antarctica 2017 campaign.

ciated with the MASC sampling area) for blowing snow and precipitation cases and compare it to existing values in the literature. Figure 15 displays the distributions of the measured size (quantified here as $D_{\max}$) for blowing snow and precipitation in Antarctica, as well as precipitation in the Swiss Alps. To help visualize the sometimes overlapping empirical distributions, the fitted gamma distributions are also plotted. The units are given in millimeters, with the approximation that 1 pixel is $\sim 33.5\,\mu m$.

As expected, the size distribution of blowing snow corresponds to smaller sizes than precipitation: the mode is around 0.2 mm for blowing snow and 0.3 to 0.4 mm for precipitation. More importantly, the right tail of the distribution is much larger for precipitation than for blowing snow. It should also be noted that the size is slightly larger in the Alpine data set (as illustrated by the slightly larger mode of the fitted gamma distributions).

Nishimura and Nemoto (2005) provide size distributions of blowing snow and precipitation measured in Antarctica at Mizuho station using a SPC. The bimodality obtained when combining blowing snow and precipitation data in Fig. 15 is in general agreement with the mixed case in their Fig. 10. However, the mode for blowing snow appears at a lower size (below 50 µm in their Fig. 7 at a height of 3.1 m). As mentioned before, this discrepancy is likely due to the limited effective resolution in MASC images after processing. In addition, as there are usually many particles in a single image dur-
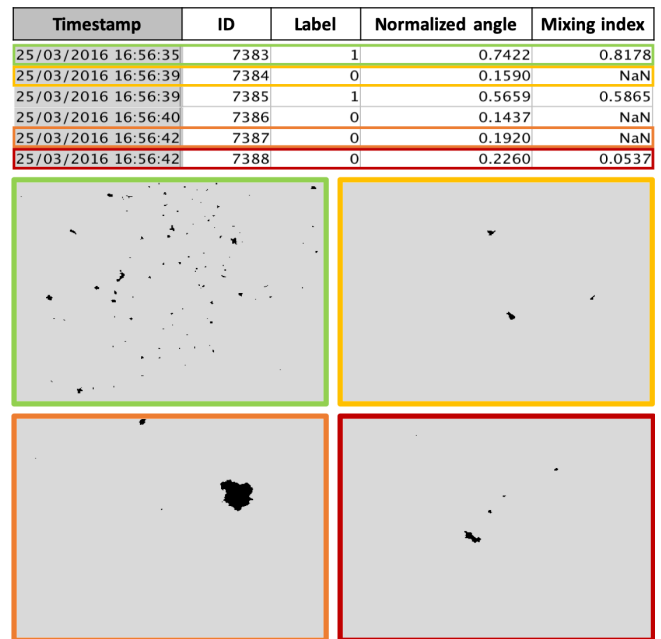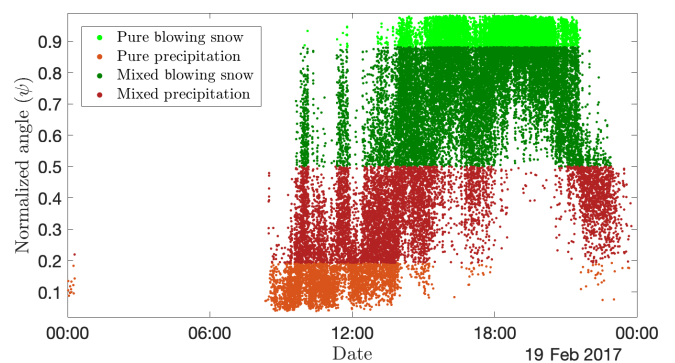
ing blowing snow, some may be out of focus and artificially appear larger than they are. So we expect the blowing snow features extracted from MASC data to be biased towards larger particles. It should also be noted that the sampling areas of the two instruments are different (see Sect. 2.1), and this could partly explain the differences in the obtained distributions.

Overall, it appears that the MASC images, processed as explained in Praz et al. (2017), are not adapted to a detailed study of the geometry of blowing snow particles but are still
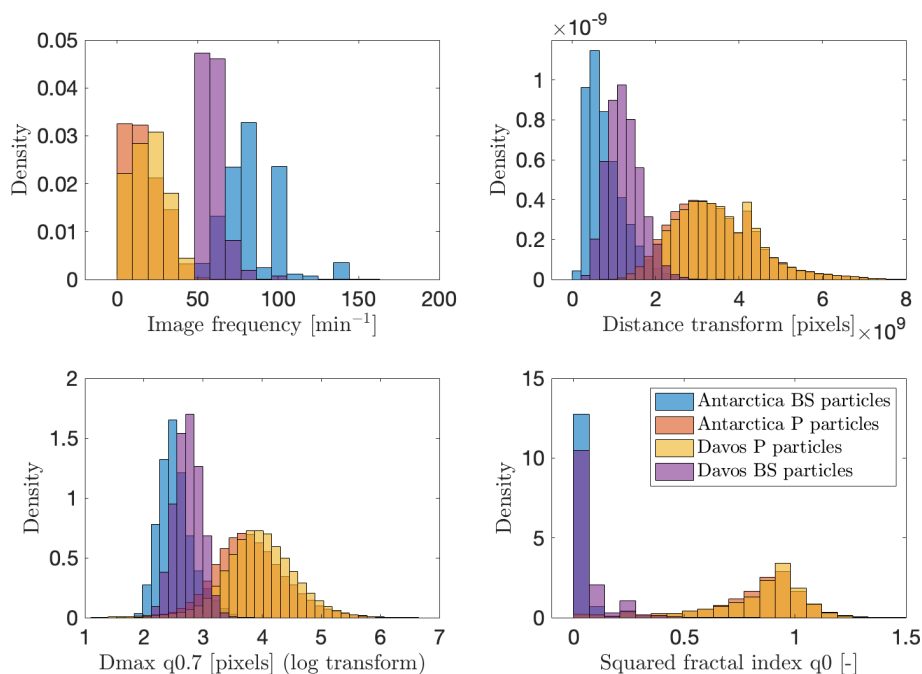
Figure 14. Histograms of selected descriptors for the training blowing snow and precipitation images from the entire Dumont d'Urville and Davos data sets.
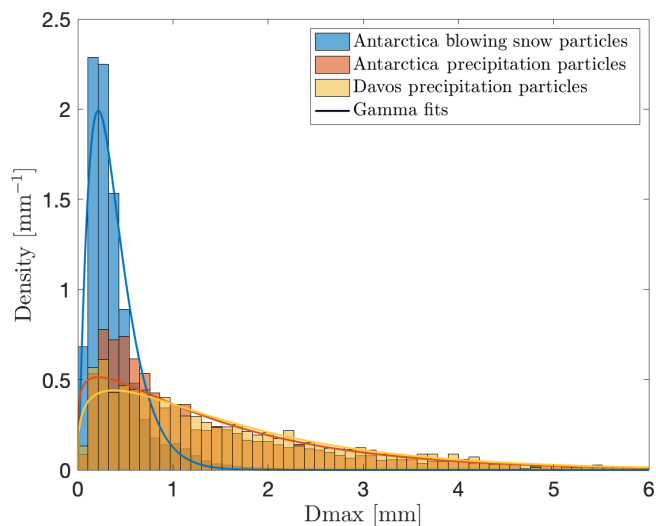

Figure 15. Histograms and fitted gamma distributions of D_{max} for images classified as pure blowing snow and pure hydrometeors from Antarctica and pure precipitation from the Alps.

relevant to distinguish blowing snow and precipitation, to characterize mixtures of both and to analyze the dynamics of blowing snow at high temporal resolutions.

6 Conclusions

A novel method to automatically detect images from the MASC instrument corresponding to blowing snow is introduced. To classify the images, the method computes four selected descriptors via image processing. The descriptors were selected to be relevant for discriminating between blowing snow particles and hydrometeors as well as to be robust to image-processing artifacts. The classification is achieved by a two-component Gaussian mixture model fitted on a subset of 8450 representative images from field campaigns in Antarctica and Davos, Switzerland. The fitted GMM is shown to reliably distinguish images corresponding to pure blowing snow and pure precipitation cases. The GMM posterior probabilities are also mapped into a new index that allows a better identification of mixed images, and a flag signals whether an image is classified as pure hydrometeor, pure blowing snow or mixed. For mixed images, an index between 0 and 1 is proposed to indicate if the image is closer to blowing snow or precipitation. Its evaluation remains qualitative as there are no quantitative observations that can be used as reference for mixed cases. The outputs are provided for each image independently or for each triplet of images (i.e., information combined over the three cameras of the MASC).

Results from a measurement campaign conducted at the Dumont d'Urville station on the coast of East Antarctica from January to July 2017 suggest that about 75 % of the images are affected by blowing snow and that about 36 % may be composed of blowing snow particles only (Table 3). The

results also suggest that about 56 % of the images could be made of a mix of blowing snow and precipitation particles, which support findings that in Antarctica blowing snow is frequently combined with precipitation (e.g., Gossart et al., 2017). Moreover, time series of the classified images highlight that blowing snow strongly relies upon fresh snow availability and often starts shortly after the beginning of precipitation (Fig. 13), which is also consistent with conclusions from Gossart et al. (2017). Results from images taken inside a double-fence intercomparison reference in Davos at 2540 m a.s.l between December 2015 and March 2016 indicate that, despite the sheltering structure, about 60 % of the images could be affected to some extent by blowing snow particles from adjacent fence ledges. In terms of the percentage of images, these numbers tend to be quite large, as the image frequency is usually much higher when strong blowing snow occurs, but the occurrence is more balanced in terms of time.

As the method was developed and tested on fundamentally different campaigns, it may have a general applicability to any other MASC images. However, it should be noted that some descriptors depend on the particular settings (e.g., image size and pixel resolution) used during the aforementioned campaigns, and a new GMM should be fitted if different settings apply. Further work should be conducted to evaluate if the method can give satisfactory results on images that do not include a timestamp, as the image frequency descriptor could not be utilized. In this case, it could be replaced by one or a couple of other descriptors listed in Table A1 of Appendix A to strengthen the model. The method could also be adjusted to train a model with a supervised learning algorithm that provides posterior probabilities such as Bayesian classifiers or logistic regression. However, this would imply some effort to increase the training set. An intercomparison between different machine-learning algorithms and the creation of different validation sets could help gain confidence in the results.

The main limitations of the present method are the assumption of normally distributed features through the use of the GMM, the too-coarse resolution of the MASC to properly capture the small end of the distribution of blowing snow particle size and the dependency of the method on the defined training set. The latter illustrates the problem of generalization. Some extremely high intensity snowfall events, higher than the ones observed during the Davos and Antarctica campaigns, could be erroneously classified as blowing snow with the current model due to the nature of the descriptors. In this case, higher-intensity pure snowfall events should be included in the training set. Another example is the size of the blowing snow particles. During the campaigns in Antarctica, the MASC was set up on a rooftop at 3 m a.g.l. Several studies have demonstrated that the size of blowing snow particles tends to decrease with height (Nishimura and Nemoto, 2005; Nishimura et al., 2014). Consequently, blowing snow particles on images from a MASC that would have been set up at much higher or lower heights may have a bias relative to the fitted Gaussian distribution of the blowing snow cluster for D_{max}. It is thus recommended to follow the procedure described in this article and fit a new model, if the one provided does not perform well in other contexts.

Appendix A: Feature extraction

Table A1. Full list of all computed descriptors. Descriptors related to each particle are transformed into a single descriptor for the image (right column). Selected ones are shown with an asterisk.

Image frequency*	–
Number of particles detected in the image	–
Distance to connect all particles	–
Number of particles smaller than a given threshold	–
Ratio of the area represented by all particles to the area of the smallest polygon encircling them	–
Cumulative distance transform*	–
Maximum diameter*	quantiles 0–1, moments 1–5
Particle area	quantiles 0–1, moments 1–5
Particle convex area	quantiles 0–1, moments 1–5
Particle perimeter	quantiles 0–1, moments 1–5
Fractal index (FRAC), fractal index squared*	quantiles 0–1, moments 1–5
Gravelius compactness coefficient (ratio of the perimeter to the one of a circle with equivalent area)	quantiles 0–1, moments 1–5

Appendix B: Image-processing issues

The median filter may not perform satisfactorily, for instance when the background luminosity is changing rapidly (see Fig. B1).

Similarly, large precipitation particles may split or appear as such in the MASC images (see Fig. B2), leading to potential biases in the number of detected particles.
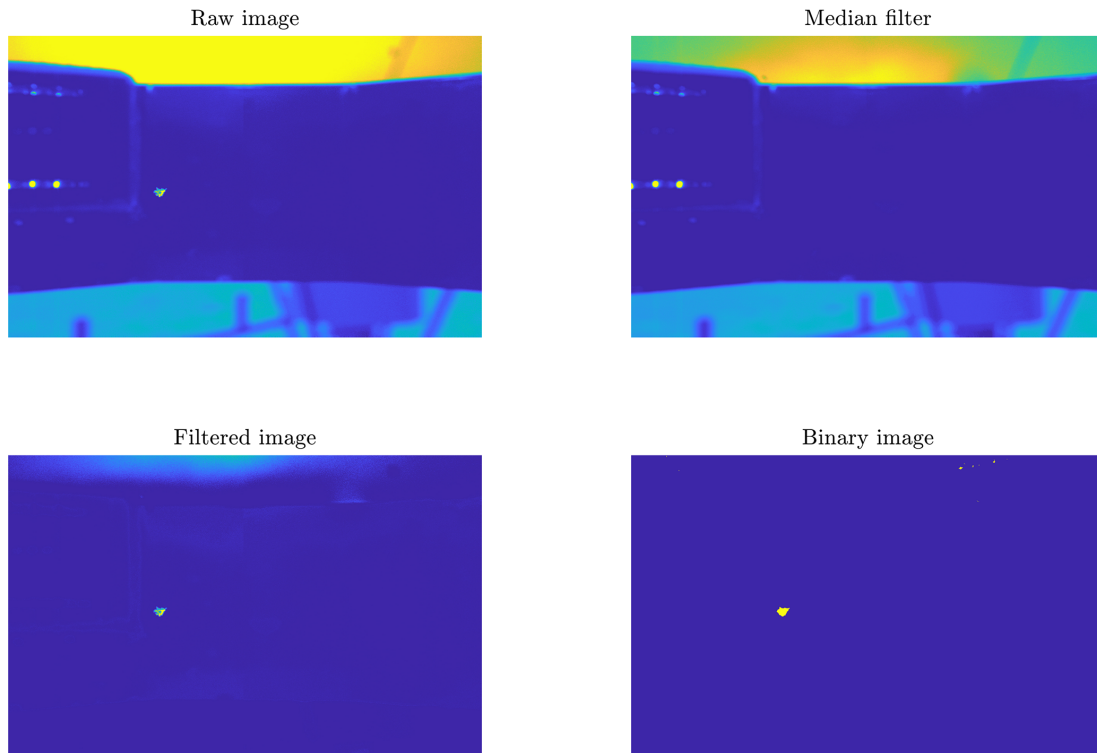
Raw image

Median filter

Filtered image

Binary image

Figure B1. Raw image, median filter, filtered image and final binary image for an example where the median filter does not perform well due to changes in sky luminosity. Some artifacts appear on the top right of the binary image
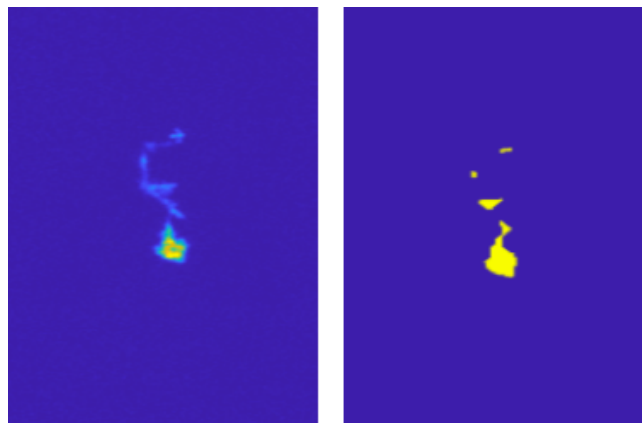
Figure B2. A precipitation particle split into fragments that could be confused with blowing snow particles. The cumulative distance transform descriptor is much less affected by such image-processing issues than the number of particles.

References

Bellot, H., Trouvilliez, A., Naaim-Bouvet, F., Genthon, C., and Gallée, H.: Present weather-sensor tests for measuring drifting snow, Ann. Glaciol., 52, 176–184, 2011.

Bintanja, R.: Snowdrift sublimation in a katabatic wind region of the Antarctic ice sheet, J. Appl. Meteorol., 40, 1952–1966, https://doi.org/10.1175/1520-0450(2001)040<1952:SSIAKW>2.0.CO;2, 2001.

Box, J. E., Bromwich, D. H., Veenhuis, B. A., Bai, L.-S., Stroeve, J. C., Rogers, J. C., Steffen, K., Haran, T., and Wang, S.-H.: Greenland ice sheet surface mass balance variability (1988–2004) from calibrated polar MM5 output, J. Climate, 19, 2783–2800, 2006.

Budd, W., Dingle, W., and Radok, U.: The Byrd snow drift project: outline and basic results, in: Studies in Antarctic Meteorology, edited by: Rubin, M. J., vol. 9 of Antarctic Research Series, American Geophysical Union, 71–134, 1966.

Budd, W. F.: The drifting of non-uniform snow particles, in: Studies in Antarctic Meteorology, edited by: Rubin, M. J., vol. 9 of Antarctic Research Series, American Geophysical Union, 59–70, 1966.

Byrt, T., Bishop, J., and Carlin, J. B.: Bias, prevalence and kappa, J. Clin. Epidemiol., 46, 423–429, 1993.

Comola, F., Kok, J. F., Gaume, J., Paterna, E., and Lehning, M.: Fragmentation of wind-blown snow crystals, Geophys. Res. Lett., 44, 4195–4203, 2017.

Crivelli, P., Paterna, E., Horender, S., and Lehning, M.: Quantifying particle numbers and mass flux in drifting snow, Bound.-Lay. Meteorol., 161, 519–542, 2016.

Das, I., Bell, R. E., Scambos, T. A., Wolovick, M., Creyts, T. T., Studinger, M., Frearson, N., Nicolas, J. P., Lenaerts, J. T. M., and van den Broeke, M. R.: Influence of persistent wind scour on the surface mass balance of Antarctica, Nat. Geosci., 6, 367–371, https://doi.org/10.1038/NGEO1766, 2013.

De Maesschalck, R., Jouan-Rimbaud, D., and Massart, D. L.: The Mahalanobis distance, Chemometr. Intell. Lab., 50, 1–18, 2000.

Déry, S. J. and Yau, M.: Large-scale mass balance effects of blowing snow and surface sublimation, J. Geophys. Res., 107, 4679, https://doi.org/10.1029/2001JD001251, 2002.

Ehbrecht, M., Schall, P., Ammer, C., and Seidel, D.: Quantifying stand structural complexity and its relationship with forest management, tree species diversity and microclimate, Agr. Forest Meteorol., 242, 1–9, 2017.

Gallée, H., Guyomarc'h, G., and Brun, E.: Impact of snow drift on the Antarctic ice sheet surface mass balance: possible sensitivity to snow-surface properties, Bound.-Lay. Meteorol., 99, 1–19, 2001.

Garrett, T. J., Fallgatter, C., Shkurko, K., and Howlett, D.: Fall speed measurement and high-resolution multi-angle photography of hydrometeors in free fall, Atmos. Meas. Tech., 5, 2625–2633, https://doi.org/10.5194/amt-5-2625-2012, 2012.

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

Genthon, C., Berne, A., Grazioli, J., Durán Alarcón, C., Praz, C., and Boudevillain, B.: Precipitation at Dumont d'Urville, Adélie Land, East Antarctica: the APRES3 field campaigns dataset, Earth Syst. Sci. Data, 10, 1605–1612, https://doi.org/10.5194/essd-10-1605-2018, 2018.

Goodison, B. E., Louie, P. Y., and Yang, D.: WMO solid precipitation measurement intercomparison, World Meteorological Organization Geneva, Switzerland, 1998.

Gordon, M. and Taylor, P. A.: Measurements of blowing snow, Part I: Particle shape, size distribution, velocity, and number flux at Churchill, Manitoba, Canada, Cold Reg. Sci. Technol., 55, 63–74, 2009.

Gossart, A., Souverijns, N., Gorodetskaya, I. V., Lhermitte, S., Lenaerts, J. T. M., Schween, J. H., Mangold, A., Laffineur, Q., and van Lipzig, N. P. M.: Blowing snow detection from ground-based ceilometers: application to East Antarctica, The Cryosphere, 11, 2755–2772, https://doi.org/10.5194/tc-11-2755-2017, 2017.

Grazioli, J., Genthon, C., Boudevillain, B., Duran-Alarcon, C., Del Guasta, M., Madeleine, J.-B., and Berne, A.: Measurements of precipitation in Dumont d'Urville, Adélie Land, East Antarctica, The Cryosphere, 11, 1797–1811, https://doi.org/10.5194/tc-11-1797-2017, 2017.

Guyomarc'h, G., Bellot, H., Vionnet, V., Naaim-Bouvet, F., Déliot, Y., Fontaine, F., Puglièse, P., Nishimura, K., Durand, Y., and Naaim, M.: A meteorological and blowing snow data set (2000–2016) from a high-elevation alpine site (Col du Lac Blanc, France, 2720 m a.s.l.), Earth Syst. Sci. Data, 11, 57–69, https://doi.org/10.5194/essd-11-57-2019, 2019.

Kennedy, P., Thurai, M., Praz, C., Bringi, V., Berne, A., and No-taroš, B. M.: Variations in Snow Crystal Riming and Z DR: A Case Analysis, J. Appl. Meteorol. Clim., 57, 695–707, 2018.

Kimura, F., Takashina, K., Tsuruoka, S., and Miyake, Y.: Modified quadratic discriminant functions and the application to Chinese character recognition, IEEE T. Pattern Anal., 1, 149–153, 1987.

Kinar, N. and Pomeroy, J.: Measurement of the physical properties of the snowpack, Rev. Geophys., 53, 481–544, 2015.

Kind, R. J.: Mechanics of aeolian transport of snow and sand, J. Wind Eng. Ind. Aerod., 36, 855–866, 1990.

Lehning, M. and Fierz, C.: Assessment of snow transport in avalanche terrain, Cold Reg. Sci. Technol., 51, 240–252, 2008.

Lenaerts, J. T. M. and van den Broeke, M. R.: Modeling drifting snow in Antarctica with a regional climate model: 2. Results, J. Geophys. Res., 117, D05109, https://doi.org/10.1029/2010JD015419, 2012.

Leonard, K. C., Tremblay, L.-B., Thom, J. E., and MacAyeal, D. R.: Drifting snow threshold measurements near McMurdo station, Antarctica: A sensor comparison study, Cold Reg. Sci. Technol., 70, 71–80, 2012.

Lesins, G., Bourdages, L., Duck, T. J., Drummond, J. R., Eloranta, E. W., and Walden, V. P.: Large surface radiative forcing from topographic blowing snow residuals measured in the High Arctic at Eureka, Atmos. Chem. Phys., 9, 1847–1862, https://doi.org/10.5194/acp-9-1847-2009, 2009.

McGarigal, K. and Marks, B. J.: Spatial pattern analysis program for quantifying landscape structure, Tech. Rep. Gen. Tech. Rep. PNW-GTR-351, US Department of Agriculture, Forest Service, Pacific Northwest Research Station, 1995.

McLachlan, G. J. and Basford, K. E.: Mixture models: Inference and applications to clustering, vol. 84, Wiley Online Library, 1988.

Moerland, P.: Mixture models for unsupervised and supervised learning, PhD thesis, available at: https://infoscience.epfl.ch/record/32665 (last access: 28 January 2020), 2000.

Mott, R. and Lehning, M.: Meteorological Modeling of Very High-Resolution Wind Fields and Snow Deposition for Mountains, J. Hydrometeorol., 11, 934–949, https://doi.org/10.1175/2010JHM1216.1, 2010.

Murphy, K. P.: Machine Learning: A Probabilistic Perspective, MIT Press, 2012.

Naaim-Bouvet, F., Bellot, H., Nishimura, K., Genthon, C., Palerme, C., Guyomarc'h, G., and Vionnet, V.: Detection of snowfall occurrence during blowing snow events using photoelectric sensors, Cold Reg. Sci. Technol., 106, 11–21, 2014.

Nishimura, K. and Nemoto, M.: Blowing snow at Mizuho station, Antarctica, Philos. T. Roy. Soc. A, 363, 1647–1662, 2005.

Nishimura, K., Yokoyama, C., Ito, Y., Nemoto, M., Naaim-Bouvet, F., Bellot, H., and Fujita, K.: Snow particle speeds in drifting snow, J. Geophys. Res., 119, 9901–9913, 2014.

Palm, S. P., Yang, Y., Spinhirne, J. D., and Marshak, A.: Satellite remote sensing of blowing snow properties over Antarctica, J. Geophys. Res., 116, D16123, https://doi.org/10.1029/2011JD015828, 2011.

Pomeroy, J. and Li, L.: Prairie and arctic areal snow cover mass balance using a blowing snow model, J. Geophys. Res., 105, 26619–26634, 2000.

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

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., Black, J., Thériault, J. M., Kucera, P., Gochis, D., Smith, C., Nitu, R., Hall, M., Ikeda, K., and Gutmann, E.: How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed, B. Am. Meteorol. Soc., 93, 811–829, https://doi.org/10.1175/BAMS-D-11-00052.1, 2012.

Scaff, L., Yang, D., Li, Y., and Mekis, E.: Inconsistency in precipitation measurements across the Alaska–Yukon border, The Cryosphere, 9, 2417–2428, https://doi.org/10.5194/tc-9-2417-2015, 2015.

Scarchilli, C., Frezzotti, M., Grigioni, P., De Silvestri, L., Agnoletto, L., and Dolci, S.: Extraordinary blowing snow transport events in East Antarctica, Clim. Dynam., 34, 1195–1206, https://doi.org/10.1007/s00382-009-0601-0, 2010.

Schmidt, R.: Threshold wind-speeds and elastic impact in snow transport, J. Glaciol., 26, 453–467, 1980.

Schweizer, J., Bruce Jamieson, J., and Schneebeli, M.: Snow avalanche formation, Rev. Geophys., 41, 1016, https://doi.org/10.1029/2002RG000123, 2003.

Shepherd, A., Ivins, E. R., Geruo, A., Barletta, V. R., Bentley, M. J., Bettadpur, S., Briggs, K. H., Bromwich, D. H., Forsberg, R., Galin, N., Horwath, M., Jacobs, S., Joughin, I., King, M. A., Lenaerts, J. T. M., Li, J., Ligtenberg, S. R. M., Luckman, A., Luthcke, S. B., McMillan, M., Meister, R., Milne, G., Mouginot, J., Muir, A., Nicolas, J. P., Paden, J., Payne, A. J., Pritchard, H., Rignot, E., Rott, H., Sandberg Sørensen, L., Scambos, T. A., Scheuchl, B., Schrama, E. J. O., Smith, B., Sundal, A. V., van Angelen, J. H., van de Berg, W. J., van den Broeke, M. R., Vaughan, D. G., Velicogna, I., Wahr, J., Whitehouse, P. L., Wingham, D. J., Yi, D., Young, D., and Zwally, H. J.: A reconciled estimate of ice-sheet mass balance, Science, 338, 1183–1189, 2012.

Vignon, É., Traullé, O., and Berne, A.: On the fine vertical structure of the low troposphere over the coastal margins of East Antarctica, Atmos. Chem. Phys., 19, 4659–4683, https://doi.org/10.5194/acp-19-4659-2019, 2019.

Winstral, A., Elder, K., and Davis, R. E.: Spatial snow modeling of wind-redistributed snow using terrain-based parameters, J. Hydrometeorol., 3, 524–538, 2002.

Yang, Y., Palm, S. P., Marshak, A., Wu, D. L., Yu, H., and Fu, Q.: First satellite- detected perturbations of outgoing longwave radiation associated with blowing snow events over Antarctica, Geophys. Res. Lett., 41, 730–735, https://doi.org/10.1002/2013GL058932, 2014.