Wet Snow Detection From Satellite SAR Images by Machine Learning With Physical Snowpack Model Labeling

Matthieu Gallet ⬩, Abdourrahmane Atto ⬩, *Senior Member, IEEE*, Fatima Karbou ⬩,
and Emmanuel Trouvé ⬩, *Senior Member, IEEE*

Abstract—The detection of wet snow by satellite imaging is currently done in an unsupervised way and lacks quantitative evaluation due to the difficulty of collecting ground truths in extreme environments. In this article, we propose to take into account information associated with a physical model to label satellite data for the purpose of supervised learning of snow properties using synthetic aperture radar (SAR) imagery. This dataset is constructed from Sentinel-1 SAR images, augmented with topographic information obtained from a digital elevation model. The labeling of this data is done at the scale of the Northern Alps using the CROCUS physical snow model. Then, we trained, over 13 combinations of labeled dataset, a wide range of machine learning models to quantitatively identify the most relevant learners for the wet snow detection task. The results demonstrate consistency among the different algorithms, with significant improvement observed when incorporating polarimetric combinations and topographic orientation data in the input of the model. The best algorithmic solution trained on this dataset is evaluated by comparing the obtained wet snow map over a validation area in the French massif of the Grandes Rousses with the existing Copernicus products, fractional snow cover, and SAR wet snow. We also compare the temporal results obtained at one meteorological station located in the test area. The results show a better representation of wet snow during the melting period using the supervised learning approach, as well as a reduction in areas classified as wet during the winter season.

Index Terms—Detection, labeling, synthetic aperture radar (SAR), segmentation, Sentinel-1, snowpack model, supervised machine learning, wet snow.

I. INTRODUCTION

THE identification of snow states, notably wet snow, has considerable significance in diverse domains owing to its multifaceted implications. To ensure effective risk management, it is essential to monitor wet snow, especially in the context of forecasting [1], avalanche awareness [2], modeling the run-out of wet snow avalanches [3], and employing physical models for avalanche simulation [4]. The impact of climate change on snow states is of significant importance. Proper assessment and understanding of these factors are necessary to mitigate the impacts of climate change. Therefore, it is crucial to evaluate its effects on communities, water resource availability [5], [6], glacier dynamics [7], [8], and hydrological basins [9]. The importance of wet snow in hydrology is critical to the hydrological cycle, influencing the timing and amount of spring snowmelt runoff [10]. The transformation of wet snow to liquid water can result in rapid and unpredictable fluctuations in river discharge, which has implications for the planning of natural resources and the management of water and energy resources [11], [12], [13]. This is especially significant within regions where snowmelt significantly contributes to water availability [14], [15]. In addition, it is important to monitor and comprehend snow and wet snow conditions to guarantee sustainable economic activities related to tourism [16], [17]. Several papers made advances in the study of the cryosphere and snow in general [18] or, more specifically, using satellite imagery [19], [20], highlighting the challenges, the prospects, and the current limitations. Satellite imagery makes it possible the global mapping, from the scale of a mountain range to a whole country. One of the largest used optical satellite indicators is the normalised difference snow index (NDSI) [21] with its combination of spectral bands, which provides information on total snow cover [22]. This index led to the creation of fractional snow cover (FSC) product of the European Copernicus program. However, the use of optical or multispectral satellites limits the analysis of the total snow cover [23], with no specific information on snow conditions: wet and dry snow. Tracking snow in the Alps via these optical satellites is difficult, especially in winter, by the presence of clouds masking portions of current acquisition or even the entire image, requiring heavy techniques for their use [24].

SAR imagery, meanwhile, with its ability to observe the Earth in all weather conditions, is a complementary solution, particularly during the winter periods. Some studies have utilized both optical and SAR data to mitigate weather disturbances [25]. Due to the ease of processing with a single sensor, SAR data has been widely used in snow studies, primarily with an X-band sensor [26] or C-band [27]. Nevertheless, the use of C-band SAR data is shown to be more appropriate for analyzing the snow properties [18]. The launch of the Sentinel-1 A and 1B C-band satellites, operating at a frequency of 5.405 GHz, allowed for continuous monitoring of Western Europe and extensive regions [28], [29], [30], [31]. These satellites provided free data

access and offered repeat pass acquisitions every 6 days using the two satellites producing, dense time series of images that are particularly interesting for regular monitoring [32], [33]. Sentinel-1B is no longer available since December 2021. Ground range detected (GRD) products are created from focused SAR data, involving detection, multilooking, and projection to ground range using an Earth ellipsoid model. The ellipsoid projection of GRD products is adjusted based on specified terrain height. The resultant product has nearly square spatial resolution and pixel spacing, with reduced speckle thanks to multilook processing. These GRD data are most often used to analyze the backscatter coefficient (σ_0) [34] in order to detect wet snow. Liquid water in wet snow induces significant dielectric losses and consequently reduce the backscatter coefficient as the liquid water content (LWC) increases. Unlike dry seasonal snow, which allows radar signals to penetrate several meters, wet snow predominantly reflects and scatters radar signals near the surface and within the uppermost layers of the snowpack [35]. This distinction results in a more notable contrast in backscatter intensity between regions with wet snow cover and those with dry or snow-free conditions [36]. However, many physical parameters, such as snow density or incidence angle, can affect this signal reduction [37]. One of the most widely employed snow detection methods relies on the approach developed by Nagler et al. [38]. In fact, Tsai et al. [19] shown that over 80% of the algorithms utilized for wet snow detection employ the backscattering signal, as in [39], [40], [41], [42]. Nagler's method based on applying a threshold on the backscatter coefficient of the ratio between a current acquisition and a snow free reference image led to the development of the SAR wet snow (SWS) Copernicus products. With the rapid growth of machine learning, some studies are exploiting radar data to adapt computer vision techniques to supervised classification problems. When analyzing the signal, one is no longer seeking a specific threshold corresponding to a physical variation. But, the emphasis is placed on recognizing attributes that can serve as distinguishing factors between classes. These features can be directly polarimetric channels [43], obtained through the use of classical textural extractors, such as GLCM[1] [44] or even through the first layers of deep learning networks as shown in the GPR[2] data in [45]. Some works have integrated the use of these supervised techniques for the study of the cryosphere, such as [46] with artificial neural networks. More specifically, a classification of wet and dry snow based on support vector machines (SVM), is proposed in [47]. Other studies have shown the benefits of using complementary data to perform wet snow classification [48], [49]. However, the main obstacle to the development and generalization of these methods over large areas is the data labeling required to train the models. In addition, the spatio-temporal nature of the data requires greater care in the process of creating and selecting training and test areas to prevent any spatial or temporal correlation [50].

Labeling is a major step in the supervised classification process but remains a challenge in the context of SAR images, especially when dealing with large areas. The main leverages for the extensive use of machine learning methods rely on this labeling, which allows to get a variety of samples that offer sufficient representativeness to enable the model to generalize the learning. Visually, the nature of SAR images complicates manual labeling. Furthermore, the ability of SAR waves to penetrate certain surfaces makes it difficult to exploit labels obtained from optical imagery. For the problem of studying snow through SAR imaging, the decrease in backscattering coefficient when the snowpack becomes wet makes the manual labeling task difficult, even impossible. Solutions have been proposed by adopting different strategies: either an unsupervised framework using an empirical threshold [35], statistical analysis like the one proposed by [51], which considers thresholding and radiative transfer modeling; or using total snow cover, obtained from MODIS [48], but incorporating a variety of SAR signal characteristics (backscattering, PolSAR, and InSAR). However, these approaches face the limits of the unsupervised methods, such as in thresholding, or when the label is too general and quite different from the radar signal sensitivity. The incorporation of topographic data is a helpful method for enhancing classification [48], [52].

In numerous studies, the final validation focuses on using weather data. However, this validation remains complex due to spatial limitations, and sometimes partially evaluated, as indicated by [19]. Nevertheless, the capability of weather models to provides detailed information at the scale of large mountain areas, as well as specific information about the nature of the studied process [53], enables fully exploiting the spatial and temporal extent of SAR data in the labeling process.

In this article, we propose a new dataset for large-scale wet snow detection, labeled from the physical model of the snowpack provided and developed by Météo-France: CROCUS [54]. The objective of the proposed approach is to study the ability of physical model to label a SAR dataset for supervised detection of wet snow in a machine learning framework. We illustrate the complete framework in Fig. 1. In Section II, we first present the data used, their characteristics, and the processing performed, followed by the labeling process with a focus on dataset structuring within the framework of supervised learning. Then, in Section III, we discuss the characteristics of the dataset created, and present the seven classic machine learning algorithms, to evaluate this dataset. In Section IV, we present a comprehensive analysis where we assess the performances of 7 algorithms and the influence of 13 data associations on the input. This evaluation involves using polarimetric, topographic, and combined polarimetric data, while quantifying the results using commonly employed metrics in machine learning. We conduct an analysis in Section IV-C with a comparison with existing Copernicus products considering the topographic characteristics of the ground and with a time series of in situ measurements from a meteorological station in the test area. Finally, Section V concludes this article.

II. LSD4WSD: WET SNOW DATASET

A. Satellite Data

SAR Data: We use GRD data obtained during the ascending orbit number 161 of Sentinel-1, denoted by A161. This orbit

---

[1]Gray Level Co-occurrence Matrice
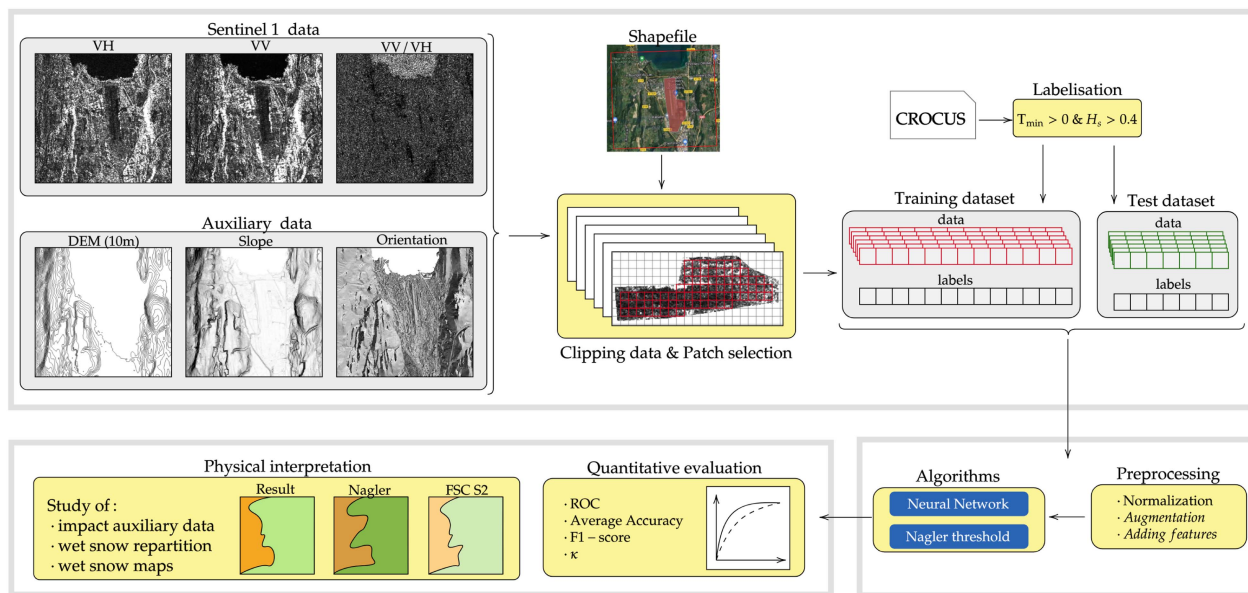[2]Ground Penetrated Radar

Fig. 1. Diagram of the complete chain for creating a labeled dataset for wet snow detection from SAR and auxiliary data and evaluating the dataset using different algorithms and metrics.
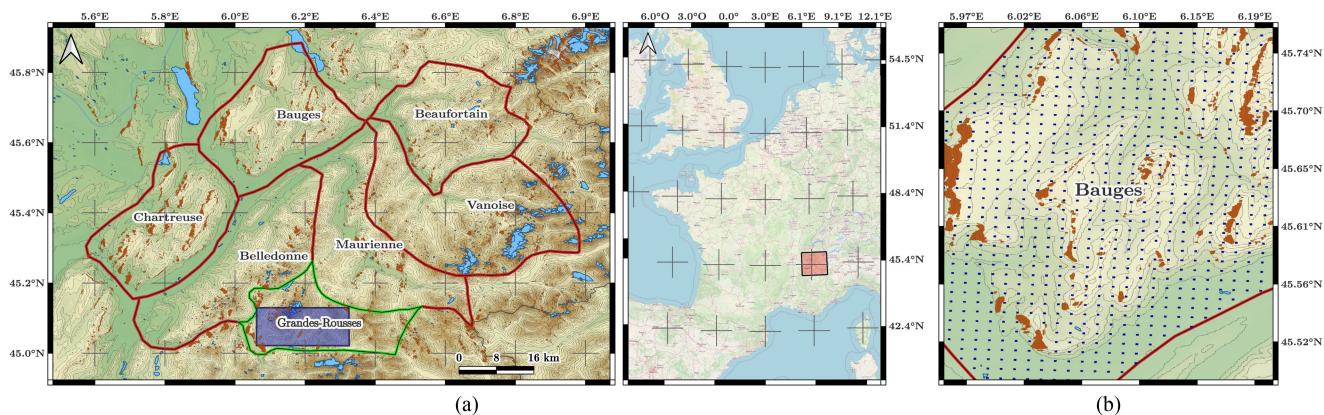

Fig. 2. (a) Location of the study area in the French Northern Alps: 6 massifs outlined in red for training and green for testing. The area inside the dark blue rectangle highlights the wet snow validation zone. In blue, the water bodies given by IGN. (b) Details of the sampling strategy in the Bauges massif for training. The small blue square represent the sample selection strategy. Layover and shadow are masked in brown.

has a local crossing time of 5:30 P.M. We use SAR images pre-processed using the CNES[3] computing facilities, which consists of a reduction of thermal noise, a calibration of the data, and a correction of the terrain using the SRTM 30 m digital elevation model (DEM). The study is conducted within a segment of the 31TGL tile, a component of both the Sentinel-2 tiling system and the military grid reference system, situated in the Northern French Alps, which covers seven French alpine massifs, as shown in Fig. 2. We finally have a series of 69 images with a temporal resolution of 6 days, using Sentinel-1 A and 1B, between August 4th 2020 and August 22nd 2021. These images are unfiltered, with a size of 10980×10980 pixels and a pixel resolution of 20 m. Each image has two polarimetric channels, VV, and VH and a third channel corresponding to the ratio of

the other two: VV/VH. Geometric distortions (layover, shadows, foreshortening) are inherent to SAR imaging. To avoid selecting these zones in the dataset, we avoid them by using a binary mask mapping the distortion zones (in orange in Fig. 2). We focused on Sentinel-1's orbit in the late afternoon to keep the geometric distortion consistent. It is worth noting that snow gets warmer after a full day, which increases LWC. By analyzing LWC at 6 P.M. and 6 A.M., we can see the difference between the descending and ascending orbit times. The median value of this difference is 0.15 kg/m^2 from October to the end of March, where we can consider a global winter-like behavior. Similarly, this difference is 1.28 kg/m^2 between April and the end of September. These values confirm a more significant presence of liquid water at the end of the day. Choosing this option helps map more expansive areas of wet snow, as the snowpack is more strongly associated with liquid water.

[3]French Space Agency.

TABLE I
COMPARISON OF THE SPATIO-TEMPORAL SPECIFICITIES OF FOUR LWC INFORMATION ACCESSES

Data	temporal resolution	data history	specificity
measuring campaign [47]	++	+	++ / +++
automatic stations [59]	++ / +++	+++	++ / +++
manual stations	++	+++	+ / ++
numerical model [54]	+++	+++	+++

(a)

Data	scale	spatial resolution	spatial density
measuring campaign [47]	local	+++	++ / +++
automatic stations [59]	global	+++	+
manual stations	global	+++	+
numerical model [54]	global	+ / ++	+++

(b)

Auxiliary Data: We use a 10 m resolution DEM provided by IGN,[4] giving access to the elevation, on which we have calculated the slope and orientation maps. We have selected an image of the GRD time series (August 9th 2020), which we consider as a reference image without snow. All these additional data aim to investigate how they are processed by the algorithms and whether or not they add any value to the quality of the classification. To compare the obtained results, we have selected two other high-level products provided by the Copernicus program. The first one is the FSC over the canopy, which gives the percentage of snow cover for each pixel with a resolution of 20 m per pixel from the Sentinel-2 data using the NDSI. The second product provided by Copernicus is the SWS obtained by the Nagler method [38] with the ascending data of Sentinel-1, with a resolution of 60 m per pixel. Because we use only ascending data, we have compared our results with SWS products obtained only in ascending passes. Two orbits covered the validation area (path 88 and path 161).

B. Labeling Data

The criteria for characterizing snow as wet are complex: LWC threshold and total or partial moistening of the snowpack. One of the main definitions is based on the LWC of the snowpack, considering the presence of wet snow from 1% of volume water content [19] or 2% of LWC for [55]. Table I summarizes four methods used to obtain the physical parameters required for labeling. These physical parameters need specific instruments (such as the Nivose stations of Météo-France), less common than the more standard instruments used to measure temperature or precipitation [56]. Compromises, therefore, have to be made. Automatic or manual measurement networks are not dense enough, so they cannot be set up in some mountainous regions to capture large variation. Measurement campaigns [57], are often limited to a small spatial and temporal extent but are highly accurate. Numerical models [58] can derive indicators directly from snowpack simulations with better temporal resolution but more limited spatial resolution.

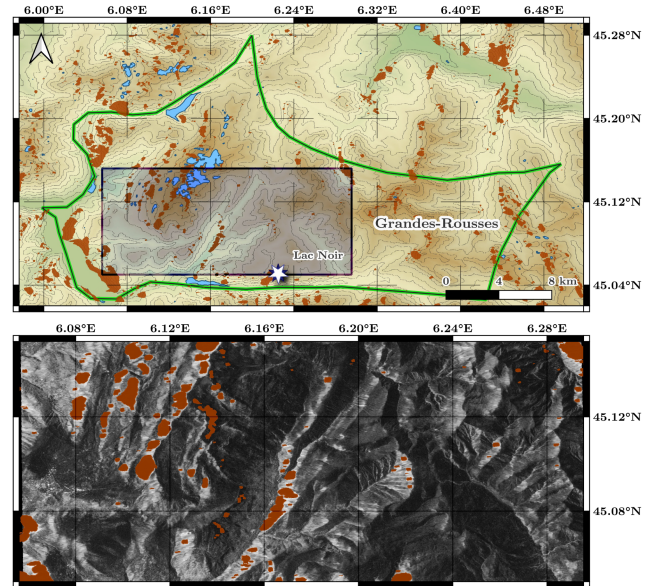[4]French National Institute for Geographic and Forestry Information

Fig. 3. Position of the validation area (semitransparent dark blue rectangle and span ($\sqrt{VV^2 + VH^2}$) of the two polarimetrics channels inside the test massif of the Grandes Rousses. The white star represents the station considered for in situ assessment. Areas affected by the SAR geometric distortions (layover and shadow) are masked in brown.

CROCUS: The Snow Study Centre (CEN) of Météo-France has developed a physically-based model CROCUS [60], which aims to describe the state and the evolution of the properties of the snowpack. The CROCUS model tracks changes in the physical properties of the snowpack, including stratigraphy (up to 50 layers) and underlying ground, in response to meteorological data reanalyzed at hourly intervals from another numerical model SAFRAN [61] but without any information of the observed snowpack. However, the primary assumption of this model is the spatial homogeneity of the considered massifs, especially for precipitation. Consequently, the spatial scale considered is that of a mountain range, but with a topographic division of 300 m in altitude (from 900 to 3600 m), for six orientations (N, E, SE, S, SW, W) and three slopes (0, 20°, 40°) and the French mountain massifs considered (including 23 Alpine massifs with a surface area of approximately 400 km^2). A graph of the temporal variation of the LWC by the model is given in Fig. 4. This scale, therefore, excludes local effects such as those due to snow accumulation or drift by the wind. CROCUS provides an output of the average snowpack described by its vertical stratigraphy and the descriptive variables of the internal snowpack, including the LWC, minimum, and maximum temperature across the snowpack and its height. To classify the type of snow, Ya-Lun et al. [19] defined wet snow based on the temperature within the snowpack. According to this definition, snow is considered wet if the temperature of the snowpack is above zero degrees Celsius. Compared to wet snow definitions based on LWC, this formulation has the advantage of remaining close to physical reality while being understandable for nonexperts with less uncertainty on the model's variable. We propose the following relation rule for wet snow labels, based on the minimum temperature ($T_{\min}$) in degrees Celsius and the
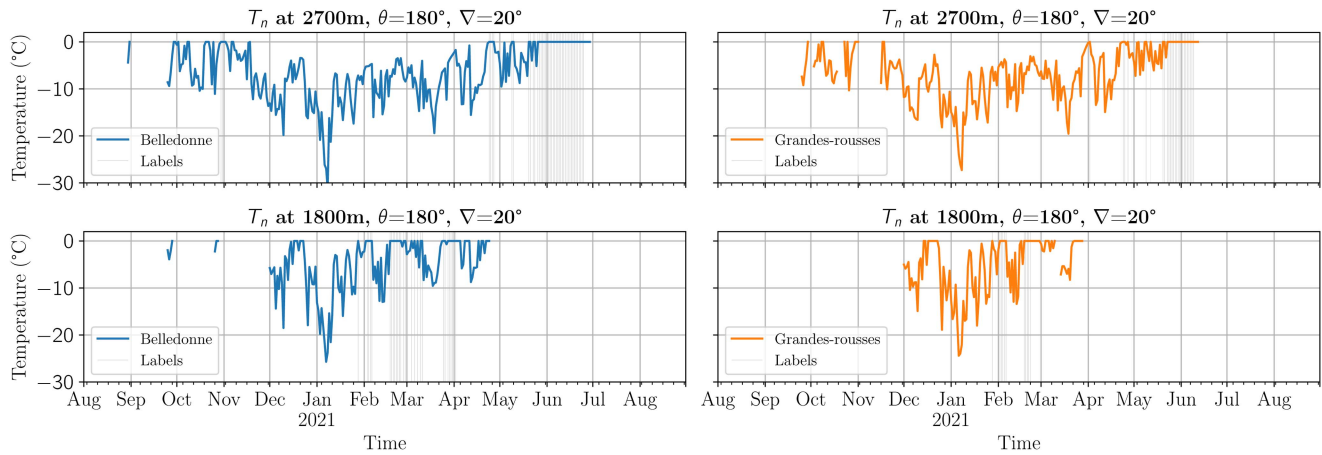
Fig. 4. Minimum snowpack temperature given by CROCUS between August 2020 and 2021. In blue, the curves are given for the Belledonne massif and in orange for the Grandes Rousses massif, for two given altitudes, facing south. The gray lines show the dates on which the (1) is respected.

snow height (H_s) in meters

$$\text{Wet Snow} = \text{T}_{\min} \geq 0 \text{ and H}_s \geq 0.4. \qquad (1)$$

By including a requirement for snow height, we can ensure that the radar signal is predominantly generated by the existence of wet snow, and not by the ground response beneath a thin layer of snow or rocky areas with limited snow coverage. This limitation also ensures to have homogeneous areas covered by snow. Fig. 4 illustrates an instance of the CROCUS output. It presents the minimum temperature of the snowpack for a southern-facing orientation and an average slope of 20° in two distinct mountain ranges: Belledonne and Grandes Rousses. These massifs are further divided into different altitudes (1500–1800 m and 2400–2700 m), representing medium and high mountain levels, respectively. Vertical gray lines on each curve represent the dates when the conditions defined by the (1) are satisfied based on the CROCUS inputs (slope, orientation, altitude, and mountain massif). Despite the geographical proximity depicted in Fig. 2, each mountain range possesses unique characteristics. The Belledone massif experiences more cases of wet snow and for longer periods throughout the season. Comparing the first and second lines in Fig. 4, the influence of the altitude gradient becomes clear. At an altitude of 1800 m, the snow is considered wet earlier in the season (February–March) but with a shorter duration of snow coverage. Given that we are working within a machine learning framework, we did not validate the samples obtained by in situ measurements. We prefer a large quantity of variable data with imperfect labeling rather than a small number of ungeneralizable ground truths. However, numerous works were able to validate the CROCUS model through measurement campaigns [60], [62], [63], compare its performance with other models [64].

C. Dataset Creation:

For a given acquisition date, we use the two polarimetric channels of Sentinel-1 and their ratio (VV/VH), three topographic channels (elevation, orientation, and slope), and the ratio of the polarimetric channels of the date with the one considered as a reference. These three additional channels of ratio with respect to the reference allow us to avoid the nonlinear calculation step that some algorithms, such as those based on decision trees, cannot perform, but also to study the decisions taken by the algorithm in the presence of only these three channels with respect to the threshold given by [65]. All channels are sampled to have a pixel size of 10 m×10 m. The pixels of the polarimetric and ratio channels are transformed in dB. All dates are used in the creation of this database. The separation of the training and test dataset is purely spatial and allows to avoid biasing the classifier by test data strongly spatially correlated to the training data. The training and test dataset creation chain is based on nine channels.

Selection of Areas: We divide the data into three categories. The training data used for learning, the test data needed for the quantitative evaluation of the model (Fig. 2 and finally the validation area used for the qualitative evaluation of the method from a spatial and temporal point of view and comparison with existing Copernicus products (Fig. 3. We have selected six massifs to built the training dataset (Beaufortain, Belledonne, Chartreuse, Maurienne, and Vanoise) and one massif for the test and validation (Grandes Rousses). The geographical distribution of these massifs is illustrated in Fig. 2. This choice was made to avoid obtaining good results due to the strong spatial correlation of the distribution of the zones. Moreover, these seven massifs are present on the 31TGL tile that we used. The selection of six massifs to train the model allows the model to be robust to the intrinsic characteristics of each massif, without resorting to domain adaptation techniques. We cut each massif using a 16×16 sliding window, and selected one window out of nine as shown in Fig. 2. The window size of 16×16 pixels was chosen empirically by experimental cross-validation study. It allows us to have enough sample in the patch to be statistically significant, its size being a multiple of 2, offers perspectives on the use of possible feature extractions (DWT,[5] CNN). This way we keep the spatial extent and diversity of the dataset, but we
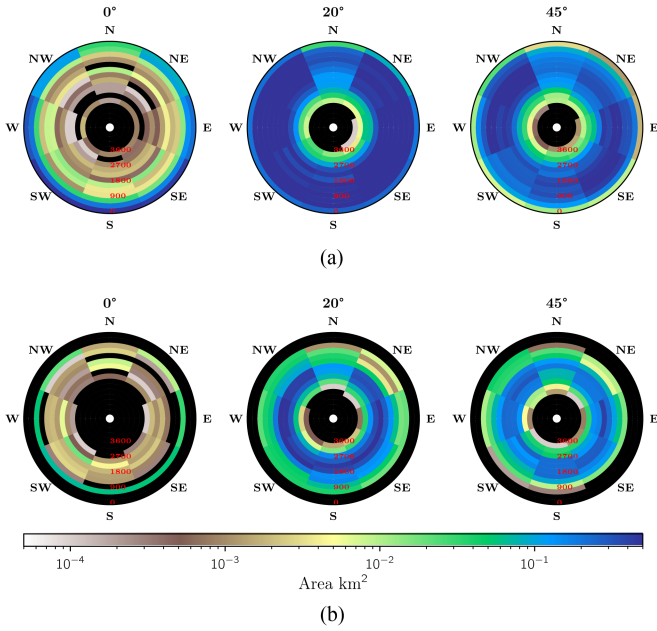
---

[5]Discrete Wavelet Transform.

Fig. 5. Topographical distributions of the test and training samples according to the discretisation given by CROCUS, for slope, orientation, and elevation. (a) Distribution of training areas for three slopes. (b) Distribution of test areas for three slopes.
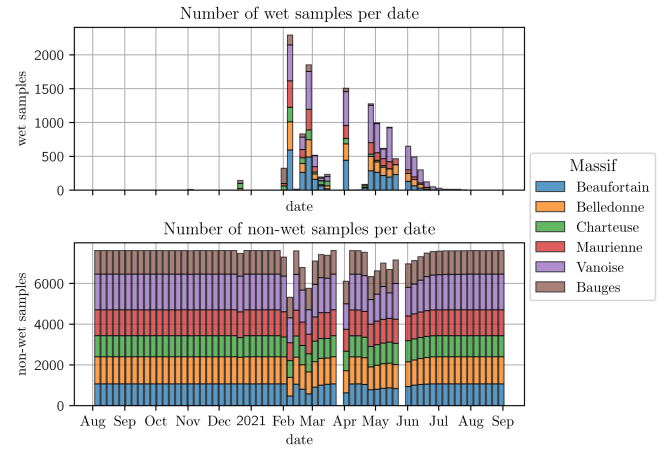

Fig. 6. Distribution of the number of samples between August 2020 and 2021, according to six training massifs considered for the two classes *wet* and *non wet*.

reduce the number of samples to reduce the computation time and reduce the spatial correlation between each sample. This paving represents, respectively, 1.4% and 1.1% of the training and test surface area of the massifs. The validation data uses the test dataset with a cutting similar to the one used for the test dataset but with a dense mesh, i.e., a stride of one instead of nine. The choice of the test massif results from the comparison made in Fig. 5.

The latter illustrates the distribution of areas in km^2 for the three topographic parameters of the CROCUS model (elevation, orientation, and slope). We note the interest of using six massifs in training by the variety of topographic characteristics that they represent together [see Fig. 5(a)]. We note that the Grandes Rousses test massif is homogeneous in terms of the distribution of these zones for all orientations and mainly for the medium to steep slopes [see Fig. 5(b)]. However, for low slopes (0°), some zones do not exist in the training dataset but appear in the test dataset. The impact of this gap is measured by the average size of these areas (of the order of 10^{-4} km^2), which is much smaller than for larger slopes (of the order of 10^{-2} km^2). Given Fig. 5(a), the training dataset is also homogeneous with ~ 10 times more surfaces for each topographical feature studied, than the selected test area. We assign the wet snow label, as defined in (1), to each patch. This label is obtained from the simulated snowpack by CROCUS at 6:00 P.M., taking into account the slope, altitude, orientation, and massif information of the respective patch.

The monthly distribution of the training dataset samples according to the wet and nonwet classes is given in Fig. 6. It is interesting to note that the nonwet class is evenly distributed across all the massifs, while the wet class is mainly represented by the Vanoise and Beaufortain massifs. The month of February

also saw a large number of wet label samples. This can be explained by the mild weather in February 2021 [66], and the presence of Saharan dust, which accelerates the warming of the snowpack [67]. This results in a classifier that performs well during this period of the year, but is less efficient in the rest of the year. The final dataset is open and available online [68].

III. Machine Learning Framework

A. Machine Learning Models: In a first step, we evaluate seven algorithms commonly used in machine learning, and select one of the best, for with the dataset. The idea is to see the performance of common solutions in this classification. Theoretical descriptions of the algorithms can be found in [69] and a brief presentation of their use in remote sensing problems [70]. The proposed selection are the following: Adaboost [71], random-forest (RF) [72], support vector machine (SVM) with Gaussian kernel [73], K-nearest neighbors (KNN) [74], logistic regression [75], multilayer perceptron (MLP) [76], and a convolutional neural network (CNN) architecture based on [77]. This choice highlights three categories of algorithms. The algorithms based on which is based on learning a set of weak classifiers (Adaboost and RF). These algorithms are well-suited to binary classification and have the advantage of allowing an interpretation of the result obtained, and thus to keep an understanding of the algorithm. The categories of algorithms are based on the concept of separation or distance. These algorithms perform the classification task by projecting the data to better distinguish classes, in other words, to maximize the separation or distance between classes (SVM). Either they directly determine the class of a new sample in relation to the distances of neighboring samples and their labels (KNN). The last category covers neural self-learning algorithms. They are based on weight learning for linear combinations and coupled with nonlinear functions. Where logistic regression learns only one level of weights, MLP is made up of a number of successive learning layers. The CNN is an extension of the MLP, using convolutions rather than direct linear combinations. Learning was achieved using the Scikit-Learn and Keras library from Python. The codes

TABLE II
PARAMETERIZATION OF THE ALGORITHMS USED

Algorithms	Settings
KNN	K neighbours = 50
SVM	kernel = Radial Basis Function (RBF)
RF	estimators = 200, criterion = entropy
AdaBoost	estimators = 200
MLP	hidden layers = 100, L2 regularisation term = 0.01
Logistic Regression	L2 penalty
CNN	3 blocks of 2D convolution + Maxpooling + ReLu

are available online: https://github.com/Matthieu-Gallet/ML-WetSnowSAR. The main parameters of the algorithms are given in Table II.

B. Preprocessing: The last step of preparation consists simply in normalizing in 0 and 1 by taking the extreme values on the set of training, test, and validation data. The minimum and maximum of all samples are used to normalize the first six channels. For the three channels built on the ratio with a snow free reference image, the extreme values used for normalization correspond to the average of the minimum and maximum of all the samples. At the end, we obtain a dataset consisting of 487157 samples of size 16 by 16 by 9 for training and 3668 for testing. From a dataset in $\mathbb{R}^{N,W_x,W_y,C}$, with N the number of samples, W_x, W_y the size of the windows and C the number of channels, we study a preprocessing step upstream of these algorithms based on two solutions:

1) The first approach consists in calculating the histogram of each channels of the data to go from $\mathbb{R}^{N,W_x,W_y,C}$ to a representation $\mathbb{R}^{N,H_{\text{bins}}}$. This processing allows the patch information to be synthesised and was used extensively in images classification [78], [79]. In the context of processing large images or large datasets, this solution allows us to keep a certain computational frugality while giving interesting performances. However, this strategy fails to take into account potential spatial structure. In order for the histogram to be statistically reliable and homogeneous, a optimal number of bins $H_{\text{bins}} = 16$ is used. This choice comes from the Freedman–Diaconis rule [80], where the optimal number is calculated with

$$H_{\text{bins}} = \left\lceil \frac{\sqrt[3]{n}(\max x - \min x)}{2 \cdot (Q_3 - Q_1)} \right\rceil \quad (2)$$

with n the number of pixels of the sample x, Q_1, and Q_3 are the first and third quantile of the sample. We use the maximum integer value of H_{bins} over the dataset and the channels. The calculation of the histogram is done on a dynamic range between 0 and 1, ensuring that all the histograms of the samples have the same support. This approach is used for the comparison of the models, except for the CNN where the images are directly given using the C channels considered, as well as for the study of the different input channels.

2) The second study focuses on the restriction of the histogram to channel statistics. Thus, it is interesting to study which statistical information takes precedence over the classification decision, and if so which statistics are

significant for distinguishing the classes. In this context, we are interested in the mean μ, standard deviation σ, skewness γ_1, and kurtosis γ_2. The latter two are defined as follows:

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}$$

$$\gamma_2 = \frac{\mu_4 - 3\mu_2^2}{\mu_2^2} \quad (3)$$

where μ_i represents the $i-$th central moment. We apply this transformation on the best algorithm and the best combination of input bands. The interest of this analysis is its ability to give explainable results, based in the KNN framework, on the Euclidean distance between a small number of variables (maximum 4) directly correlated to the nature of the signal.

The test dataset is balanced to have an equal number of samples from the wet and nonwet classes. However, balancing the training dataset by undersampling would result in a significant proportion of samples being removed and hence reducing the variability of the dataset (the wet class being in the minority with 13975 samples, we would reduce the size of the dataset of nearly 95%). To address this issue, we propose a variation of the K-Fold technique [81]: the balanced stratified K-Fold. We divide the majority class of the dataset into K sub-datasets of a size equal to the number of samples in the minority class. Each subdataset of the majority class is combined with the minority class to form a fold. This way, we can sweep across the entire dataset and study the variability of the models across all the balanced folds. In our case, we obtain $K = 34$ fold with 13975 samples of each class.

C. Evaluation: Quantitative Analysis: In order to evaluate the performance of each method, associated with transformations or auxiliary data, we first focus on a quantitative analysis. We choose four metrics to characterize each model with T_P, T_N the true positives and negatives, F_P, F_N the false positives and negatives, and N the number of samples considered

1) Accuracy $= (T_P + T_N)/N$.
2) F1-score $= T_P/[T_P + 0.5(F_N + F_P)]$.
3) Kappa [82].
4) Receiver operating characteristic (ROC) curves.

The first three metrics are used to analyze the results of the classification of wet snow using all algorithms and only the polarimetrics channels VV and VH. We then use specifically the F1-score to study the sensitivity of the four best algorithms previously tested with respect to various channel combinations. All the bands combinations are presented in Table III. We have selected 13 different band combinations. The ones specific to the polarimetric information of the SAR signal can be grouped together ($\mathbf{A}, \mathbf{B}, \mathbf{H}, \mathbf{H_r}$). Note that the $\mathbf{G}$ combination is simply the Nagler case when a fixed thresholding classifier is used directly. This band is therefore important for the comparison of the performances with respect to the other combinations. We also consider the groupings made up of the VV and VH polarimetric channels augmented with topological information ($\mathbf{C}, \mathbf{D}, \mathbf{E}, \mathbf{F}$). The idea is to see if the topological information is decisive in

TABLE III
DESCRIPTION OF THE TESTED CHANNEL COMBINATIONS

Identifier	Channel combinations
A	VV, VH
B	A, VV/VH
C	A, elevation
D	A, orientation
E	A, slope
F	A, elevation, orientation, slope
G	VV_{ratio}, VH_{ratio}
H	A, VV_{ratio}, VH_{ratio}
H_r	A, VV/VH, VV_{ratio}, VH_{ratio}, VV_{ratio}/VH_{ratio}
I	A, VV_{ratio}, VH_{ratio}, orientation
I_D	A, VV_{ratio}, VH_{ratio}, elevation
I_P	A, VV_{ratio}, VH_{ratio}, slope
J	F, G, VV/VH, VV_{ratio}/VH_{ratio}

TABLE IV
RESULTS FOR 3 METRICS FOR THE SEVEN CLASSIFIERS CONSIDERED FOR THE SNOW DETECTION PROBLEM USING K-FOLD

	Adaboost	RF	KNN	CNN
Accuracy	70.9 ± 0.3	71.8 ± 0.3	$\mathbf{73.2 \pm 0.3}$	71.5 ± 1.9
F1-score	74.7 ± 0.2	75.4 ± 0.2	$\mathbf{75.9 \pm 0.3}$	70.7 ± 2.4
Kappa	41.8 ± 0.7	43.6 ± 0.6	$\mathbf{46.5 \pm 0.6}$	43.1 ± 3.7

	Logistic Regression	MLP	SVM
Accuracy	70.4 ± 0.2	72.0 ± 2.6	72.6 ± 0.2
F1-score	74.5 ± 0.1	75.0 ± 1.0	75.6 ± 0.1
Kappa	40.8 ± 0.4	43.9 ± 5.1	45.1 ± 0.3

The bold values indicate the best score for earch metrics considered.

addition to the backscatter signal to differentiate the classes. The counterpart of this grouping is made up of the remaining classes, and the contribution of the ratio of a sample at a current date to a reference date is studied in addition.

We select four different input combinations to plot ROC curves. On the ROC curves, we study more particularly the thresholding giving the maximum precision (BAROC) and the thresholding giving a false positive rate lower than 5% (FCROC). These values are obtained by computing the ROC curve from the probabilities provided by the classifier. We then recover the thresholding that maximizes the accuracy and the one that gives a false positive rate of 5%. In this way, we measure what can be the maximum performances of the model, but also the performances with a low false positive rate compensating a poor or noisy labeling.

Qualitative Analysis: In a second step, we are interested in the qualitative evaluation of the methods and their comparison with existing products (FSC and SWS). The FSC product is produced using optical satellite data from the Sentinel-2 (revisit time of 5 days). This product provides the percentage of the surface covered by snow at the top of the canopy (FSC-TOC) per pixel, with a spatial resolution of 20 m. To ensure accuracy, atmospheric correction, and cloud masking are utilized. The snow-covered area is identified using a thresholding method based on the NDSI and a digital elevation model. The fractional snow cover of the snow-covered pixels is then calculated using an empirical relationship with NDSI that has been calibrated with higher resolution satellite images [22].

The SWS product is based on Sentinel-1 satellites. To map the extent of snowmelt areas, the process depends on change detection, which involves comparing the reduced backscatter coefficient of wet snow to conditions where the surface is free of snow or covered by dry snow. Specifically, this product is based on Nagler's method [65], and uses S1 Interferometric Wide Swath (IW) mode with dual-pol (VV, VH) acquisitions. The resulting product offers wet snow extent data for high mountain regions, with a spatial resolution of 60 m. To ensure accuracy, the product masks radar shadow, layover, and foreshortening, as well as water bodies, forests, urban areas, and nonmountain regions.

These evaluations are done on the specific area of the validation dataset (Grandes-Rousses massif) as illustrated in Fig. 3. We utilize a set of two distinct dates that are carefully selected to ensure that both the SWS and FSC products are available within 2 days of the target date. Furthermore, these dates are chosen to give an observation of the classification results:

1) during snowmelt, when wet snow represents a significant part of the snowpack,
2) and in the middle of winter, when the snow cover is important but mainly made up of fresh or dry snow.

On each of the dates, we use two spatial tools:

1) Geographical maps of wet snow probability. The resulting maps are the probabilistic output of the classifier P_{wet} (the probability of wet snow). This information is discretized into three levels: $P_{wet} > 0.5$, $P_{wet} > $ FCROC, and $P_{wet} > $ BAROC. The BAROC and FCROC thresholds are determined from the training dataset and are the average of the thresholds on each fold. This gives us the information of the wet snow extension if the maximum accuracy is considered, if a lower false alarm rate of 5% is considered, and finally the classical binary output of a classifier.

2) Topographic diagram. It gives the percentage distribution of wet pixels for a slope and orientation slice as a function of altitude. It allows us to compare the quality of classification with respect to the products and its sensitivity according to the topological characteristics.

IV. PERFORMANCE EVALUATION AND MACHINE LEARNING PRODUCTION OF WET SNOW MAPS

A. Evaluation of Models

The results of the evaluation of the seven models for the 34 folds with the two bands VV and VH are given in Table IV. We can see that the best results for the three metrics are obtained with the KNN model. In our case, we use this model with $K = 50$, the number of neighbors. It is interesting to note that the results of the KNN model are better than the results of the CNN model, although the CNN model is a deep learning model, which is more complex than the KNN model.

In addition, the variability of the two model of the family of deep learning models (CNN and MLP) is higher than the other models. This factor is quite important, because it shows the overfitting of some of the models. This can be explained by
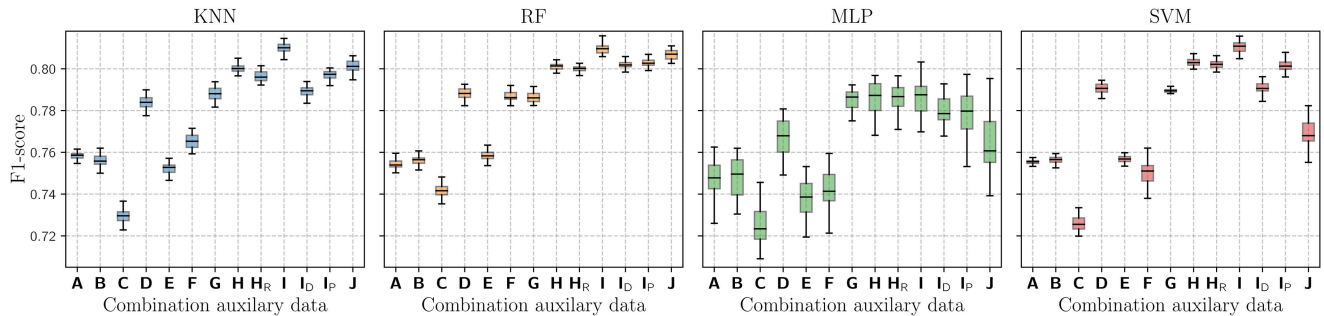
Fig. 7. Evaluation of the F1-score for the four best algorithms obtained in Table IV as a function of the combinations of the considered channels.

the fact that each of the folds, an instance of a classifier is trained and then tested on the same dataset. If the variability across folds is large, this means that each instance of the classifier is more concerned with maximizing the metrics specific to the training dataset than with finding a sufficiently general model. This reduction in variability goes hand in hand with the generalization of the model over all folds, but leads to a slight decrease in some criteria such as accuracy compared to other models such as the KNN. In terms of needs, the two best classifiers in this study are the KNN, for its overall average performance, which is better than the other classifiers. But also the SVM, which is close to the KNN metrics with a good stability on all folds. Given these results, we chose to select the four best algorithms in terms of F1-score (SVM, RF, MLP, and KNN) to evaluate the performance of theses models as a function of the combinations of the input bands (see Table III). This selection details the behavior of the three main proposed families. More results can be found in [83], where the authors focus specifically on the RF and CNN studies.

B. Evaluation of Auxiliary Data

The results are given in Fig. 7. For the KNN algorithm, it can be noted that the solution using the VV and VH channels (A) alone gives a baseline for the other combinations with a value around 0.76 and a low variability. The addition of the VV/VH ratio (B) slightly degrades the performance. The most striking feature is the addition of topographical information. The elevation in particular does not allow to refine the results, on the contrary to the orientation, which improves the F1-score by more than 2%. It can be seen that the combination (G) only composed of the VV and VH ratios of the current dates with respect to the reference date gives better results than the combinations using both the simple polarimetric channels and the topographical information. The attempt to use all the channels to have the best possible performance is not the best solution, but rather the H and I solutions, which offer a compromise between the best performance and a lower variability over all the K-Folds. It is observed that the performance of the considered algorithms is fairly consistent across the channels examined. However, the MLP approach exhibits a significant degree of variability when applied to combinations. The RF algorithm is noteworthy for its good performance at large and minimal variability across all

folds. This methodology has been previously utilized in [48]. For the remainder of this study, our focus will be on the KNN algorithm due to its ability to incorporate a distance measure between samples. Additional analyses can be found in [83] on RF and CNN.

Fig. 8 gives the ROC curves for a selection of combinations for the wet class. On the left figure, BAROC ($\times$) and FCROC ($+$) thresholds have been added. These thresholds were calculated from the training data. This explains why the FCROC thresholds, calculated for a false positive rate of 5%, are shifted. The best classifiers on this criterion are those with the lowest false positive test rate of 5%. The G combination performs best on this criterion with a false positive test rate of 8% against $\sim 10\%$ and 16% for the others. On the criterion of the area under the curve, the two combinations I and J give the best performances in terms of detection, with a slight advantage for combination I, which gives its maximum accuracy with a false positive rate lower than that of combination J.

These results can be linked to the right part of the Fig. 8, where we studied the ability of the mean backscattering of the VV and VH channels to detect the wet snow class. To do this, we used the training dataset and calculated the average value of each sample for VV and VH. We notice that both thresholds are obtained for a lower true positive rate than the methods proposed in the right part of the figure, demonstrating the benefit of both the additional channels and a classification based on a more complex algorithm than a thresholding.

On Nagler's Thresholding: The thresholding algorithm can then be considered and the ROC curve and the BAROC and FCROC thresholds associated with this mean can be calculated. The FCROC value gives a threshold value between 0 and 1 for which only 5% of false positives are obtained (and similarly for the BAROC value). Thus, from these threshold values, and knowing the dynamics used to normalize the data, we can arrive at a threshold value in dB that can be compared to that traditionally used by the Nagler method. The results obtained are given in Table V. It can be seen that the traditional thresholding value of -3 dB is close to the threshold obtained with the FCROC thresholding. The thresholding based on the average of a 16×16 pixels window, to obtain a false positive rate of the wet snow class compared to the labeling rule given in 1 is similar to the threshold proposed in [35]. This result validates the proposed labeling based on CROCUS. We also notice that when
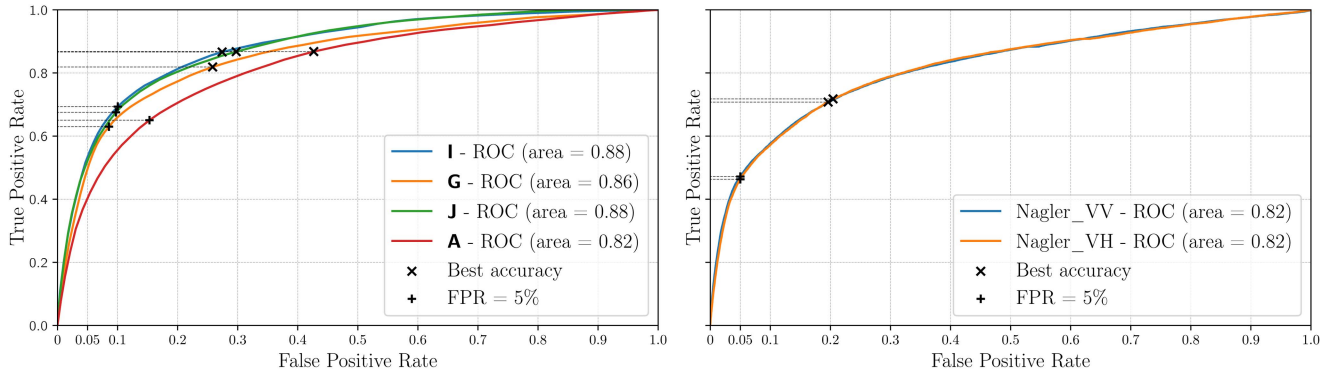
Fig. 8. ROC curve for four input channel combinations for the KNN classifier and their associated BAROC and FCROC metric at 5% (left). ROC curve for the simple thresholding type classifier on the mean of each sample for the VV and VH channels (Nagler thresholding), with BAROC and FCROC metrics (right).

TABLE V
THRESHOLD VALUE FOR BAROC AND FCROC METRICS FOR LEARNING ON
THE SAMPLE MEAN FOR VV AND VH

	VV	VH
min/max (dB)	-35.96 / 36.42	-33.30 / 29.53
BAROC (dB)	-0.99	-1.33
FCROC 5% (dB)	-2.61	-3.40

TABLE VI
DESCRIPTION OF THE TESTED STATISTICAL COMBINATIONS, WHERE μ, σ ARE
THE MEAN AND STANDARD DEVIATION AND γ_1, γ_2 ARE GIVEN IN (3)

Identifier	Statistics combinations
M	μ
N	μ, σ
O	μ, σ, γ_1
P	$\mu, \sigma, \gamma_1, \gamma_2$

we select the BAROC thresholding which gives the thresholding for the maximum accuracy, the detection threshold tends to be higher than -3 dB. Consequently, we can deduce that in the framework of this labeling, the optimal threshold to detect samples respecting the proposed labeling is about 2 dB higher than the Nagler threshold. This results in the detection of a larger area. This conclusion is based on a learning of the threshold respecting the CROCUS labeling from ascending data.

C. Statistics Set

We study the performances of the KNN classifier, when using statistics as presented in Section III-A. We study four combinations described in Table VI, with μ the mean, σ the standard deviation, and γ_1, γ_2, respectively, the skew and kurtosis. These statistics are calculated on the *G*-band combination, in order to observe results that depend only on polarimetric data. The results obtained are given in Fig. 9.

We notice at first that the use of the mean of each of the four channels (*L*) considered gives a difference of more than 14% on the F1-score. However, we reduced the number of features for each sample by 16 to perform this task with only 4 features. The addition of second-order moments improves the performance. The addition of the third-order moment (*N*)
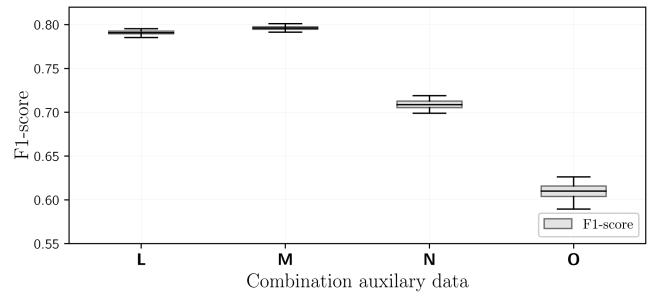

Fig. 9. F1-score for the KNN classifier as a function of the combination of statistics described in Table VI on the **G** channel combinations.

reduces the variability of the score over all the folds. However, the *O* combination has a lower F1 score than *N*, and *M* and shows increased variability, and the addition of the fourth moment *P* brings the f1 score down to 60%. The best score is obtained using the mean and standard deviation combination (*N*) of each channel. It is interesting to note that with only the first two moments calculated for each channel, we obtain $\sim 99\%$ of the F1-score obtained using the full histogram on each channel. We observe that moments of order higher than two reduce the F1-score drastically. This may be due to two factors. The first is the imprecision that increases when trying to estimate higher order moments. The second is due to the nature of the distributions of the two classes, where the Gaussian character takes precedence, giving rise to superior performance only thanks to the first two moments.

Fig. 10 illustrates the distributions of each of the six possible combinations for the four channels as a function of the mean and standard deviation, i.e., the 2-D histogram for the sample mean [see Fig. 10(a)] and the sample standard deviation [see Fig. 10(b)].

It can be seen that the distributions associated with the mean of the samples are more distinct for the two classes than those associated with the standard deviation. However, the latter show significant differences in shape (R_VV versus VV) and even in mode (R_VV versus R_VH). The distributions related to the mean are more different between the two classes: while the dry class appears to be strongly Gaussian, which can be explained
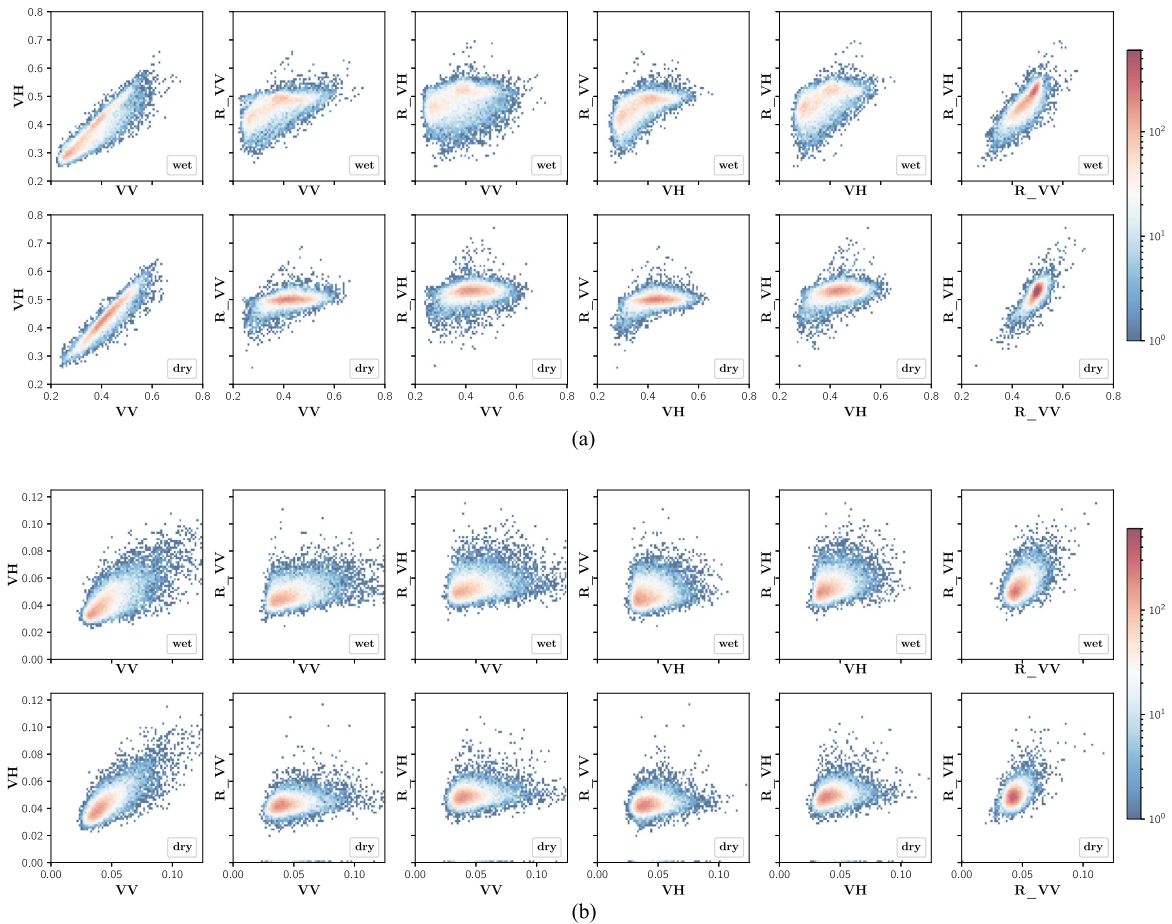
Fig. 10. Comparison of 2-D histogram for both mean and standard deviation of each samples, for the two classes *wet* and *not wet*, and between the different channels for (VV, VH R_VV and R_VH). R_VV and R_VH represent the ratio between the current acquisition and the reference image described in I, respectively, in VV and VH polarization. (a) Comparison of the mean for each samples of the four considered channels. The first line of plots corresponds to the *wet* class and the second to the *nonwet* class. (b) Comparison of the standard deviation for each samples of the four considered channels. The first line of plots corresponds to the *wet* class and the second to the *nonwet* class.

by the averaging of 256 of medium resolution, that of the wet class is more complicated.

There are marked differences in the mode of some of the distributions (R_VV versus R_VH), but above all we can observe a bimodal aspect, especially in the distributions between the primary polarimetric channels (VV,VH) and the ratios. We analyzed the composition of the modes present in the distribution formed by the VH and R_VV channels. We selected 5000 samples of the wet class around the two main modes M_{b1} and M_{b2} whose barycenters are located, respectively, around $(x_{M_{b1}}, y_{M_{b1}}) = (\sim 0.3, \sim 0.42)$ and $(x_{M_{b2}}, y_{M_{b2}}) = (\sim 0.4, \sim 0.48)$. The first difference between the samples is the acquisition date. While the samples composing the first mode M_{b1} are homogeneously distributed between February and June, the samples of the second mode M_{b2} are mainly concentrated on February. The second distinction between the two modes is the geographical position of the samples. We note that the samples belonging to the M_{b1} mode are mainly located in rocky areas, on plains or bare ground. On the other hand, the samples from the M_{b2} mode are mainly located in forest areas. It is interesting to note that the position of M_{b2} is very close to the single mode for the nonwet

class samples. This is due to the fact that in the presence of a drill the wave penetration is too weak to reliably distinguish the presence of wet snow, at least from a point of the signal average on the resolution cell. However, the presence of such samples may allow the use of the histogram to differentiate certain areas with low forest cover.

D. Focus on Specific Situations

The study area on which the wet snow map is made is shown in Fig. 3. Two acquisition dates are considered: January 18, 2021 and March 31, 2021. The Copernicus SWS products are obtained for these same dates, while the optical FSC products are obtained for January 19, 2021 and March 30, 2021, respectively. To perform this study, we use the results obtained by the KNN classifier using the histograms on each of the bands of the *I* combination.

Wet Snow Map: Based on this combination, the training dataset allows us to calculate the BAROC and FCROC thresholds. Traditionally, classifier outputs are thresholded at 0.5, with higher values considered as belonging to the class under

consideration. It is proposed to add the two additional thresholds BAROC and FCROC. The maximum accuracy is reached at 79.7% with a BAROC threshold of 0.48. We obtained an accuracy of 75.2% with a constant false positive rate of 5%, for an FCROC threshold of 0.72. The classification results are performed using a sliding window. The classification value of the central pixel is, respectively, 0,1,2, or 3 if there is no wet snow, wet snow obtained by BAROC thresholding, wet snow by classical thresholding at 0.5, and wet snow by FCROC thresholding. The results are shown in Fig. 11(a).

We first notice is that the addition of the BAROC threshold does not bring much to the results obtained with the KNN compared to the usual thresholding (0.5). However, we can see the advantage of using the thresholds together, with higher resulting pixel values in areas where the threshold above which the training classifier made few misclassification.

We need to be careful about directly comparing results on January 18, 2021. Indeed on this date, the CROCUS model classify $\sim 5\%$ of the pixels in this area as wet snow according to the proposed criterion. The FSC product does not detail the difference between wet and dry snow and gives the overall snow cover result, while the SWS product, because of its threshold, tends to overestimate our criterion.

On the date of 31 March 2021, this time we are in the melting period and we can see a more marked correlation between the three results. Where the SWS product gives large homogeneous areas, we find a more important clipping, as in the extreme right part of the image, where we find the five diagonal bands corresponding to bowl faces. This clipping has the opposite effect of obtaining more heterogeneous maps.

Fig. 11(b) represents a zoom in Fig. 11(a) on the Alps d'Huez for March 31st 2021. For SWS products, we have added the product mask categories. In this way, the SWS product masks both city and forest areas. Interestingly, the KNN result offers a finer segmentation of wet snow. In addition, based on the variety of samples, we can see that city areas are correctly taken into account compared to the FSC optical product, as are forest areas.

Wet Snow Topographic Repartition: We then evaluated the distribution of pixels considered as wet snow for the topographical parameters of orientation and altitude, for the two dates (31 March 2021 and 18 January 2021), but over the whole Grandes Rousses massif (given in green in Fig. 2). We considered a step of 100 m for the altitude and $12.5°$ for the orientation. To simplify the analysis we did not consider the slope variable, and calculated this distribution on all slopes below $45°$. This evaluation was carried out on the same results as in Section IV-D, i.e., KNN with the bands combination I. We studied separately the results obtained by the FCROC thresholding, the BAROC one, and compared with the distribution calculated for the SWS products. We also compared the snow cover distribution of the FSC product. The representation in Fig. 12 gives in percentage for each cell over the whole area considered without clouds for the FSC and without geometric distortion.

For the date of January 18, 2021, we find that the three results based on the SAR and therefore sensitive to wet snow, are less distributed than the true snow extension. It can be seen that the FCROC, BAROC and SWS give a maximum of wet snow for elevations between 1800 m and 2400 m facing East.

For the date of 31 March 2021, we notice that the BAROC threshold diagram is closer to that of the FSC than the SWS, notably in terms of percentage, we find the two main modes for elevations between 1800 and 2400 m with an East and West orientation, respectively, and this in a marked way in the two diagrams, whereas the SWS is slightly more homogeneous on all orientations.

A number of pixels are detected as wet snow for elevations below 1800 m for orientations between West and South West, but only in the FCROC and BAROC results. These areas correspond to the detection of two lakes in the southern part of the Grandes Rousses massif.

E. In-Situ Comparison

We consider the Lac Noir weather station, positioned as shown in Fig. 3, providing the following daily in-situ measurements: snow depth (cm), minimum daily temperature (degrees Celsius), and precipitation (mm). To make the figures easier to read, we have normalised precipitation between 0 and 1, where 1 is the maximum precipitation during the period considered (corresponding to a precipitation of 23.5 mm). Fig. 13(a) and (b) shows the temporal variation of wet snow prediction by the KNN method, as a function of the three in situ information for two combinations of channels (A, I). The temporal window is between January 2020 and early June 2020. The prediction is made on a 2 by 2 pixel window centered on the station. We have added to the figures the dates on which the Copernicus SWS product indicated the presence of wet snow. It is interesting to note that the A combination gives fairly homogeneous results with no particular trends. If we set the probability threshold at 0.5, we find the information given by the SWS product. It is difficult to see a clear correlation between these data and the in situ measurements. From April onward, we notice that a trend is emerging and the number of dates and the associated probability of wet snow increases. This is explained by the progressive increase in the temperature of the snowpack and the start of snowmelt. Fig. 13(b) gives the variation for the combination giving the best results in the tests proposed in the manuscript. There are clearly two regimes, one before mid-March when the probability is low even at times when there is a sudden increase in temperature. The other regime, after mid-March, shows an increase in the probability of detection, correlated with the permanent increase in temperature. This probability is higher than that given by combination A. In all scenarios, the drop in the probability of wet snow at the beginning of June is explained by a significant reduction in the snowpack, at the limit or even below the labeling threshold (40 cm).

V. DISCUSSIONS

It is important to note that the information used for labeling is imperfect. However, this allows us to add information on a large number of samples and to use the diversity of the samples to guide sufficiently generalized learning for the evaluation of our problem. We note that the maps created using the KNN algorithm have a higher degree of heterogeneity than the FSC and SWS products. However, the resolution of SWS in particular is 60 m per pixel, whereas we propose maps with 10 m per pixel.
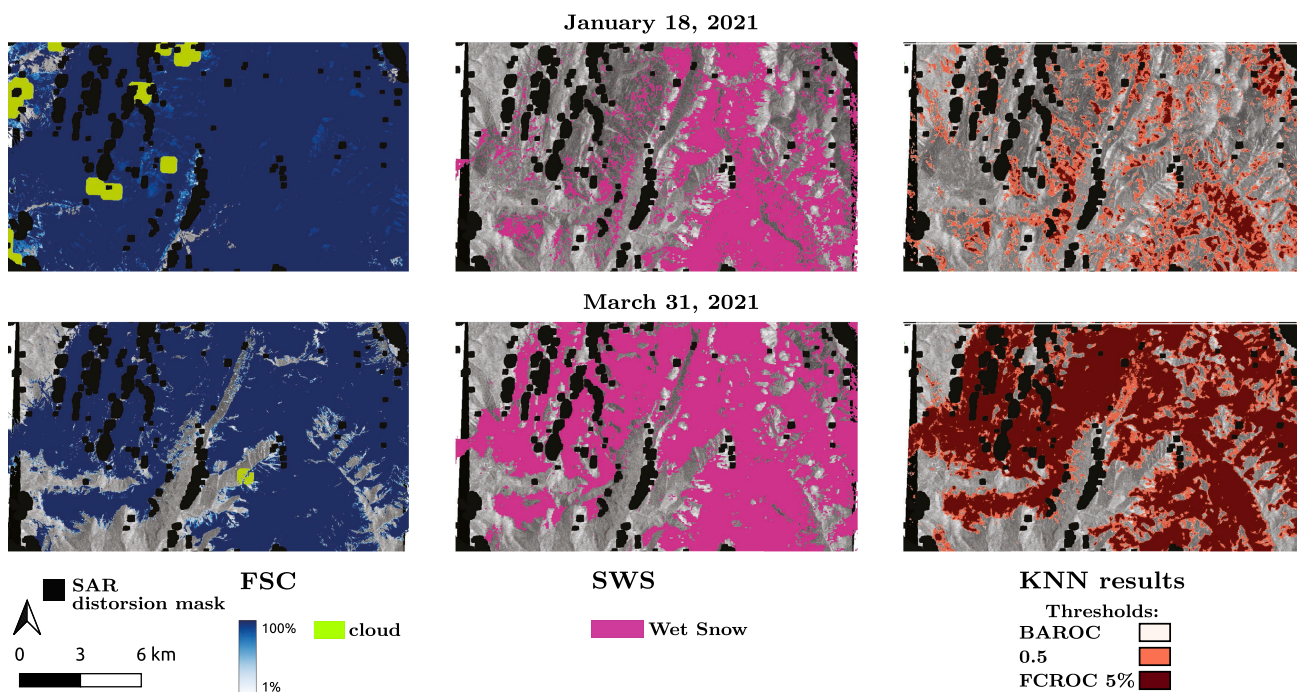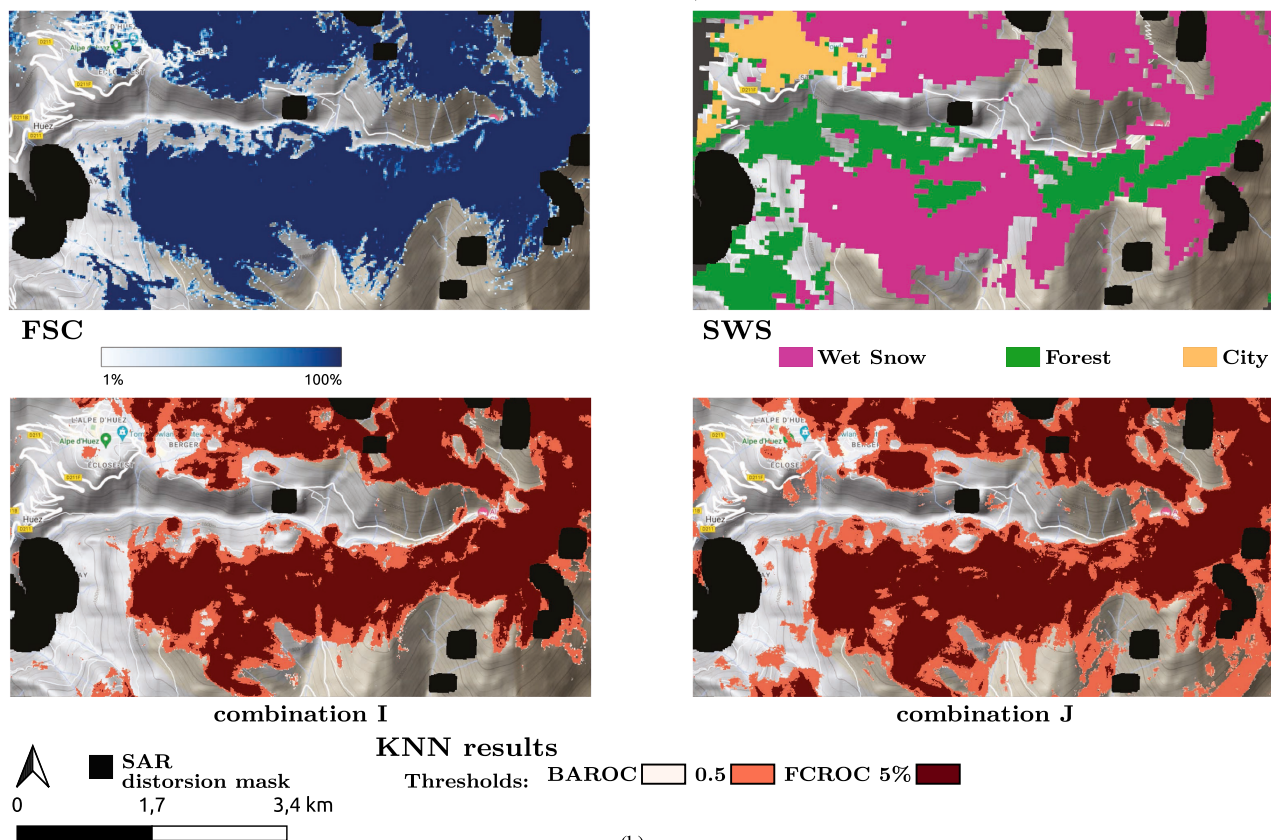
Fig. 11. Wet snow maps for different solutions using KNN, compared with existing products. (a) Maps of snow classification results for two dates in comparison with Copernicus FSC and SWS products. (b) Map details and zoomed-in view of snow classification results for the two best combinations *I* and *J* comparing Copernicus FSC and SWS products in the Alpe d'Huez, France during the melting period March 31st 2021.
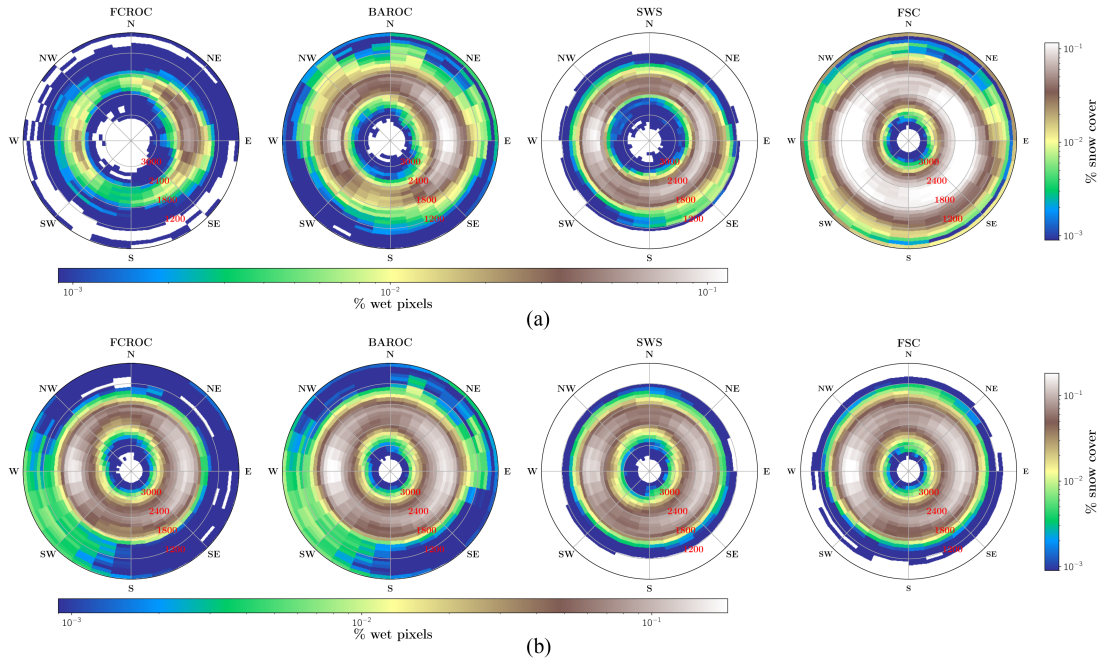
Fig. 12.   Diagrams of the distribution of wet pixels (FRCOC, BAROC, SWS) and snow cover (FSC) as a function of altitude and orientation for a slope less than 45°, for two dates. (a) January 18, 2021. (b) March 31, 2021.
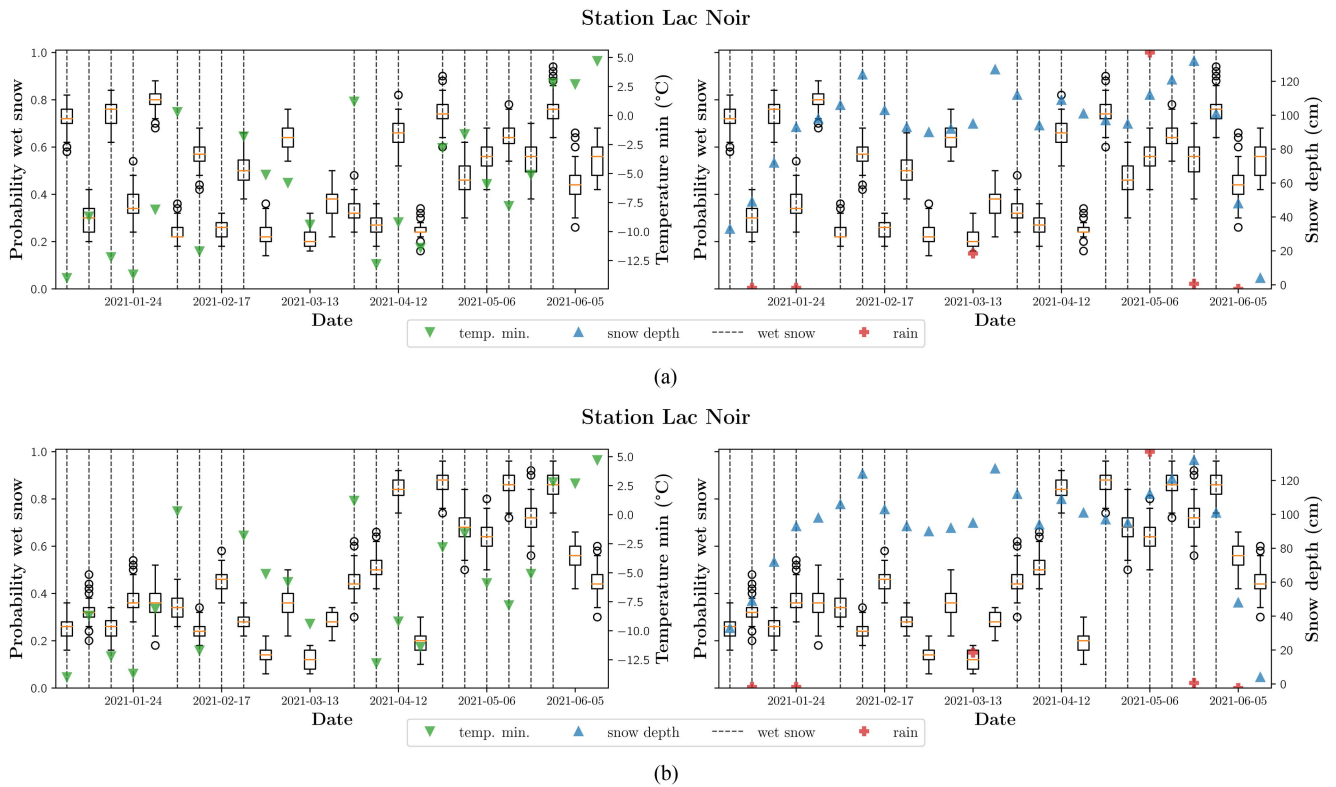

Fig. 13.   Comparison of Lac Noir station measurements with KNN's prediction. (a) KNN's prediction using *A* channel combinations. (b) KNN's prediction using *I* channel combinations.

It may be interesting to carry out an analysis in further work on the scores and classification quality as a function of the type of soil. It can be seen that the solution using all the channels (*I*) tends to overestimate the areas considered as wet snow. This is clearly seen in Fig. 11(b), where areas that are not even detected by the FSC product as having snow are detected as wet snow.

The results presented in Fig. 13(b) are interesting because the Copernicus method clearly shows that snow is detected early in the season where, for a snow depth of more than 40 cm, the minimum temperature is low, mostly below 0 or even $-2.5\,°C$, whereas our proposals give low probability rates. This can be explained by the impact of frost on the radar signal. General studies have shown the sensitivity of the C-band radar signal to the detection of frozen ground [84], causing a drop in the backscattered signal in a similar way to the consequence of the presence of wet snow [85]. We can hypothesise that the thresholding method for Copernicus products is sensitive to this frost and detects it as wet snow, whereas our method, by learning from a large number of features associated with a rich training data, does not consider these areas as wet snow.

VI. CONCLUSION

In this work, we have proposed a new dataset for the problem of detecting wet snow with SAR images, based on the use of a physical model of the snowpack covering the entire Alps. We successfully demonstrated the effectiveness of machine learning algorithms in classifying wet snow. Our approach involved using a novel labeling method based on the physical model of the snowpack CROCUS. We tested seven classifiers of a different kind after implementing our new labeling framework. KNN was the most effective. We analyzed input channels, including histograms and topographical information, to inspect our classification methods. We found that reducing the channels to statistics of order 1–4 improved our understanding of the classification decisions made by the classifier. We compare the results at the scale of the Grandes Rousses massif with the main reference snow products.

Interestingly, beyond the numerical performance in Section IV, which notably outperforms the fixed threshold approach, the classification results show coherent maps with the existing products to be put in perspective with the proposed label. Testing different classifiers helps us analyse the performance of the training data, which is imbalanced because of the object's temporal nature. This can be seen when evaluating and comparing the results with existing Copernicus products. There has been a clear improvement in detection sharpness, although this may tend to reveal heterogeneous detection zones. The size and variety of the data used also makes it possible to dispense with zone masks, which can be complex to manage in the case of simple thresholds such as forest or city zones. A possible improvement would be to restrict the temporality of the dataset to the melting season, by increasing the spatial density, to have a classification valid only for some months of the year when wet snow can be found (this excludes cases of rain on snow events). A second possible improvement would be the aggregation or fusion of the different results in a supervised or unsupervised ensemble approach. We have observed that the SWS maps and the outcomes produced by our classifier, along with a blend of different bands, yield comparable results. However, each method possesses unique specificities that distinguish them from one another. This specificity is generalized if we look at the results of other classifiers or other band combinations. Thus, the aggregation of these model outputs according to the same criterion or another more general one would allow us to obtain more detailed maps.

REFERENCES

[1] S. Bellaire, A. van Herwijnen, C. Mitterer, and J. Schweizer, "On forecasting wet-snow avalanche activity using simulated snow cover data," *Cold Regions Sci. Technol.*, vol. 144, pp. 28–38, 2017.

[2] S. Baggi and J. Schweizer, "Characteristics of wet-snow avalanche activity: 20 years of observations from a high alpine valley (dischma, Switzerland)," *Natural Hazards*, vol. 50, pp. 97–108, 2008.

[3] C. Vera Valero, N. Wever, M. Christen, and P. Bartelt, "Modeling the influence of snow cover temperature and water content on wet-snow avalanche runout," *Natural Hazards Earth Syst. Sci.*, vol. 18, no. 3, pp. 869–887, 2018.

[4] N. Wever, C. Valero, and C. Fierz, "Assessing wet snow avalanche activity using detailed physics based snowpack simulations: Simulating wet snow avalanche activity," *Geophys. Res. Lett.*, vol. 43, pp. 5732–5740, 2016.

[5] A. Bahrami, K. Goïta, and R. Magagi, "Analysing the contribution of snow water equivalent to the terrestrial water storage over Canada," *Hydrological Processes*, vol. 34, no. 2, pp. 175–188, 2020.

[6] H. Castebrunet, N. Eckert, and G. Giraud, "Snow and weather climatic control on snow avalanche occurrence fluctuations over 50 yr in the french alps," *Climate Past*, vol. 8, no. 2, pp. 855–875, 2012.

[7] A. Snehmani, P. K. Bhardwaj, A. J. Pandit, and A. Oanju, "Study of temporal changes in snout position and wet snow line for gangotri glacier using remote sensing, ground observations and meteorological data," *Int. J. Geoinformatics*, vol. 9, no. 1, pp. 49–60, May 2013.

[8] S. Adam, A. Pietroniro, and M. M. Brugman, "Glacier snow line mapping using ERS-1 SAR imagery," *Remote Sens. Environ.*, vol. 61, no. 1, pp. 46–54, 1997.

[9] B. Snapir, A. Momblanch, S. K. Jain, T. W. Waine, and I. P. Holman, "A method for monthly mapping of wet and dry snow using sentinel-1 and modis: Application to a himalayan river basin," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 74, pp. 222–230, 2019.

[10] R. Caves, O. Turpin, T. Nagler, and D. Miller, "The role of earth observation in snowmelt runoff monitoring from high latitude basins: Sar aspects," in *Proc. Sens. Manag. Environ. IEEE Int. Geosci. Remote Sens. Symp.*, 1998, pp. 1858–1860.

[11] C. Viel, A.-L. Beaulant, J.-M. Soubeyroux, and J.-P. Céron, "How seasonal forecast could help a decision maker: An example of climate service for water resource management," *Adv. Sci. Res.*, vol. 13, pp. 51–55, 2016.

[12] F. Habets, E. Philippe, E. Martin, C. H. David, and F. Leseur, "Small farm dams: Impact on river flows and sustainability in a context of climate change," *Hydrol. Earth Syst. Sci.*, vol. 18, no. 10, pp. 4207–4222, 2014.

[13] M. Beniston, "Climatic change in mountain regions: A review of possible impacts," *Climatic Change*, vol. 59, pp. 5–31, 2003.

[14] G. Rondeau-Genesse, M. Trudel, and R. Leconte, "Monitoring snow wetness in an alpine basin using combined C-band SAR and modis data," *Remote Sens. Environ.*, vol. 183, pp. 304–317, 2016.

[15] L. Braun, E. Brun, Y. Durand, E. Martin, and P. Tourasse, "Simulation of discharge using different methods of meteorological data distribution, basin discretization and snow modelling," *Nordic Hydrol.*, vol. 25, pp. 129–144, 1994.

[16] H. François, S. Morin, M. Lafaysse, and E. George-Marcelpoil, "Crossing numerical simulations of snow conditions with a spatially-resolved socio-economic database of ski resorts: A proof of concept in the french alps," *Cold Regions Sci. Technol.*, vol. 108, pp. 98–112, 2014.

[17] P. Spandre et al., "Climate controls on snow reliability in French Alps ski resorts," *Sci. Rep.*, vol. 9, pp. 1–9, 2019.

[18] L. S. Taylor, D. J. Quincey, M. W. Smith, C. A. Baumhoer, M. McMillan, and D. T. Mansell, "Remote sensing of the mountain cryosphere: Current capabilities and future opportunities for research," *Prog. Phys. Geography: Earth Environ.*, vol. 45, no. 6, pp. 931–964, doi: 10.1177/03091333211023690.

[19] S. Ya-Lun, A. Tsai, N. D. Oppelt, and C. Kuenzer, "Remote sensing of snow cover using spaceborne SAR: A review," *Remote Sens.*, vol. 11, no. 12, 2019, Art. no. 1456.

[20] A. W. Nolin, "Recent advances in remote sensing of seasonal snow," *J. Glaciol.*, vol. 56, no. 200, pp. 1141–1150, 2010.

[21] F. R. Valovcin, "Snow/cloud discrimination," 1976. [Online]. Available: https://api.semanticscholar.org/CorpusID:127774553

[22] S. Gascoin, M. Grizonnet, M. Bouchet, G. Salgues, and O. Hagolle, "Theia snow collection: High-resolution operational snow cover maps from sentinel-2 and landsat-8 data," *Earth Syst. Sci. Data*, vol. 11, no. 2, pp. 493–514, 2019.

[23] F. Koch et al., "Retrieval of snow water equivalent, liquid water content, and snow height of dry and wet snow by combining GPS signal attenuation and time delay," *Water Resour. Res.*, vol. 55, no. 5, pp. 4465–4487, 2019.

[24] D. K. Hall, G. A. Riggs, N. E. DiGirolamo, and M. O. Román, "Modis cloud-gap filled snow-cover products: Advantages and uncertainties," *Hydrol. Earth Syst. Sci. Discuss*, vol. 123, pp. 1–23, 2019.

[25] G. He et al., "Dry and wet snow cover mapping in mountain areas using SAR and optical remote sensing data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2575–2588, Jun. 2017.

[26] S. Leinss, G. Parrella, and I. Hajnsek, "Snow height determination by polarimetric phase differences in X-band SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 9, pp. 3794–3810, Sep. 2014.

[27] J. T. Koskinen, J. T. Pulliainen, K. P. Luojus, and M. Takala, "Monitoring of snow-cover properties during the spring melting period in forested areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 50–58, Jan. 2010.

[28] H. Lievens et al., "Snow depth variability in the northern hemisphere mountains observed from space," *Nature Commun.*, vol. 10, no. 1, Art. no. 4629.

[29] F. Karbou et al., "Monitoring wet snow over an alpine region using sentinel-1 observations," *Remote Sens.*, vol. 13, 2021, Art. no. 381.

[30] P. Thakur et al., "Estimation of snow cover area, snow physical properties and glacier classification in parts of western himalayas using C-band SAR data," *J. Indian Soc. Remote Sens.*, vol. 45, pp. 525–539, 2016, doi: 10.1007/s12524-016-0609-y.

[31] C. Liu, Z. Li, P. Zhang, L. Huang, Z. Li, and S. Gao, "Wet snow detection using dual-polarized sentinel-1 SAR time series data considering different land categories," *Geocarto Int.*, vol. 37, no. 25, pp. 10907–10924, 2022.

[32] N. Longépé, S. Allain, and E. Pottier, "Snow wetness monitoring using multi-temporal polarimetric ASAR data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2007, pp. 2644–2647.

[33] G. Beltramone, M. Scavuzzo, A. German, and A. Ferral, "Wet snow detection in patagonian andes with Sentinel-1 SAR temporal series analysis in GEE," in *Proc. IEEE Congreso Bienal de Argentina*, 2020, pp. 1–8.

[34] F. Karbou, I. Gouttevin, and P. Durand, "Spatial and temporal variability of wet snow in the French mountains using a color-space based segmentation technique on Sentinel-1 SAR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 5586–5588.

[35] T. Nagler and H. Rott, "Retrieval of wet snow by means of multitemporal SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 754–765, Mar. 2000.

[36] P. Mann, "Spatial and temporal variability of the snow environment in the western canadian arctic," 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:133921856

[37] C. Marin et al., "Use of sentinel-1 radar observations to evaluate snowmelt dynamics in alpine regions," *Cryosphere*, vol. 14, no. 3, pp. 935–956, 2020.

[38] T. Nagler and H. Rott, "Retrieval of wet snow by means of multitemporal SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 754–765, Mar. 2000.

[39] N. Besic, G. Vasile, J.-P. Dedieu, J. Chanussot, and S. Stankovic, "Stochastic approach in wet snow detection using multitemporal SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 244–248, Feb. 2015.

[40] A. Guiot, F. Karbou, G. James, and P. Durand, "Insights into segmentation methods applied to remote sensing SAR images for wet snow detection," *Geosciences*, vol. 13, no. 7, p. 193, 2023.

[41] F. Karbou et al., "Monitoring wet snow over an alpine region using sentinel-1 observations," *Remote Sens.*, vol. 13, no. 381, pp. 2575–2588, 2021.

[42] G. He et al., "Dry and wet snow cover mapping in mountain areas using SAR and optical remote sensing data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2575–2588, Jun. 2017.

[43] L. Huang, Z. Li, B.-S. Tian, Q. Jiu-Liang, C. Liu, and R. Zhang, "Classification and snow line detection for glacial areas using the polarimetric SAR image," *Remote Sens. Environ.*, vol. 115, no. 7, pp. 1721–1732, 2011.

[44] Y. Wang, Q. Yu, W. Lv, and W. Yu, "Coastline detection in SAR images using multi-feature and SVM," in *Proc. 4th Int. Congr. Image Signal Process.*, 2011, pp. 1227–1230.

[45] M. Gallet, A. Mian, G. Ginolhac, and N. Stelzenmuller, "Classification of GPR signals via covariance pooling on CNN features within a riemannian framework," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2022, pp. 365–368.

[46] G. Palermo, E. Raparelli, P. Tuccella, M. Orlandi, and F. S. Marzano, "Using artificial neural networks to couple satellite c-band synthetic aperture radar interferometry and alpine3D numerical model for the estimation of snow cover extent, height, and density," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 2868–2888, 2023.

[47] C. Liu et al., "Identifying wet and dry snow with dual-polarized c-band SAR data based on Markov random field model," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 2000305.

[48] O. D. Tsai and Kuenzer, "Wet and dry snow detection using sentinel-1 SAR data for mountainous areas with a machine learning technique," *Remote Sens.*, vol. 11, no. 8, 2019, Art. no. 895.

[49] R. Nijhawan, J. Das, and B. Raman, "A hybrid of deep learning and hand-crafted features based approach for snow cover mapping," *Int. J. Remote Sens.*, vol. 40, no. 2, pp. 759–773, 2019.

[50] R. Hänsch, "The trap of random sampling and how to avoid it - alternative sampling strategies for a realistic estimate of the generalization error in remote sensing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 2020–2023.

[51] N. Besic, G. Vasile, J.-P. Dedieu, J. Chanussot, and S. Stankovic, "Stochastic approach in wet snow detection using multitemporal SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 244–248, Feb. 2015.

[52] C. Liu, Z. Li, P. Zhang, and Z. Wu, "Seasonal snow cover classification based on SAR imagery and topographic data," *Remote Sens. Lett.*, vol. 13, no. 3, pp. 269–278, 2022.

[53] M. Yari et al., "Airborne snow radar data simulation with deep learning and physics-driven methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 12035–12047, 2021.

[54] E. Brun, P. David, M. Sudul, and G. Brunot, "A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting," *J. Glaciol.*, vol. 38, no. 128, pp. 13–22, 1992.

[55] F. Techel and C. Pielmeier, "Point observations of liquid water content in wet snow - investigating methodical, spatial and temporal aspects," *Cryosphere*, vol. 5, pp. 405–418, 2010.

[56] A. Lundberg, N. Granlund, and D. Gustafsson, "Towards automated 'ground truth' snow measurements–a review of operational and new measurement methods for Sweden, Norway, and Finland," *Hydrological Processes*, vol. 24, no. 14, pp. 1955–1970, 2010.

[57] J. Odry, M. A. Boucher, P. Cantet, S. Lachance-Cloutier, R. Turcotte, and P. Y. St-Louis, "Using artificial neural networks to estimate snow water equivalent from snow depth," *J./Revue Canadienne Des Ressources Hydriques*, vol. 45, no. 3, pp. 252–268, 2020.

[58] S. Yamaguchi, A. Sato, and M. Lehning, "Application of the numerical snowpack model (snowpack) to the wet-snow region in Japan," *Ann. Glaciol.*, vol. 38, pp. 266–272, 2004.

[59] A. Muhuri, S. Manickam, A. Bhattacharya, and Snehmani, "Snow cover mapping using polarization fraction variation with temporal radarsat-2 C-band full-polarimetric SAR data over the indian Himalayas," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 7, pp. 2192–2209, Jul. 2018.

[60] V. Vionnet et al., "The detailed snowpack scheme crocus and its implementation in SURFEX v7," *Geoscientific Model Develop. Discuss.*, vol. 4, 2011, doi: 10.5194/gmdd-4-2365-2011.

[61] Y. Durand, E. Brun, L. Merindol, G. Guyomarc'h, B. Lesaffre, and E. Martin, "A meteorological estimation of relevant parameters for snow models," *Ann. Glaciol.*, vol. 18, pp. 65–71, 1993.

[62] F. Larue et al., "Simulation and assimilation of passive microwave data using a snowpack model coupled to a calibrated radiative transfer model over northeastern Canada," *Water Resour. Res.*, vol. 54, no. 7, pp. 4823–4848, 2018.

[63] M. Lehning, P. Bartelt, B. Brown, T. Russi, U. Stöckli, and M. Zimmerli, "Snowpack model calculations for avalanche warning based upon a new network of weather and snow stations," *Cold Regions Sci. Technol.*, vol. 30, no. 1/3, pp. 145–157, 1999.

[64] L. H. Richard, E. Essery, H. Martin, A. C. D. Fernandez, and E. Brun, "A comparison of four snow models using observations from an alpine site," *Climate Dyn.*, vol. 15, pp. 583–593, 1999.

[65] T. Nagler, H. Rott, E. Ripper, G. Bippus, and M. Hetzenecker, "Advancements for snowmelt monitoring by means of sentinel-1 SAR," *Remote Sens.*, vol. 8, no. 4, 2016, Art. no. 348.

[66] M. France, "Hiver 2020-2021 : Très arrosé avec une alternance de temps hivernal et de douceur printanière (Hiver météorologique: décembre - janvier - février) (bilan définitif au 9 mars 2021)," 2021. Accessed: Dec. 12, 2023. [Online]. Available: https://meteofrance.fr/sites/meteofrance.fr/files/files/editorial/bilan_hiver_definitif_2020-2021_090321.pdf

[67] O. Dick et al., "Can saharan dust deposition impact snowpack stability in the french alps?," *Cryosphere*, vol. 17, no. 4, pp. 1755–1773, 2023.

[68] M. Gallet, A. Atto, F. Karbou, and E. Trouvé, "LSD4WSD: An open dataset for wet snow detection with SAR data and physical labelling," 2023. [Online]. Available: https://doi.org/10.5281/zenodo.8111485

[69] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Berlin, Germany: Springer, 2009.

[70] D. J. Lary, H. A. Alavi, H. A. Gandomi, and A. L. Walker, "Machine learning in geosciences and remote sensing," *Geosci. Front.*, vol. 7, no. 1, pp. 3–10, 2016.

[71] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Proc. Eur. Conf. Comput. Learn. Theory*, 1995, pp. 23–37.

[72] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.

[73] C. Cortes and V. N. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.

[74] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.

[75] J. A. D. Anderson, "Separate sample logistic discrimination," *Biometrika*, vol. 59, pp. 19–35, 1972.

[76] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychol. Rev.*, vol. 65, no. 6, 1958, Art. no. 386.

[77] H. Parikh, S. Patel, and V. Patel, "Classification of SAR and PolSAR images using deep learning: A review," *Int. J. Image Data Fusion*, vol. 11, no. 1, pp. 1–32, 2020.

[78] D. Dai, W. Yang, and H. Sun, "Multilevel local pattern histogram for SAR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 225–229, Mar. 2011.

[79] K. Kayabol, "Histogram-based contextual classification of SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 33–37, Jan. 2015.

[80] A. D. Freedman and P. Diaconis, "On the histogram as a density estimator:l2 theory," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 57, pp. 453–476, 1981.

[81] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statist. Surv.*, vol. 4, pp. 40–79, 2010.

[82] J R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–74, 1977.

[83] M. Gallet, A. Atto, E. Trouvé, and F. Karbou, "CNN classification of wet snow by physical snowpack model labeling," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2023, pp. 5351–5354.

[84] T. Jagdhuber, J. Stockamp, I. Hajnsek, and R. Ludwig, "Identification of soil freezing and thawing states using SAR polarimetry at C-band," *Remote Sens.*, vol. 6, no. 3, pp. 2008–2023, 2014.

[85] N. Baghdadi, H. Bazzi, M. E. L. Hajj, and M. Zribi, "Detection of frozen soil using sentinel-1 SAR data," *Remote Sens.*, vol. 10, no. 8, 2018, Art. no. 1182.

Abdourrahmane Atto (Senior Member, IEEE) is currently a Professor with the University Savoie Mont Blanc, Polytech Annecy-Chambéry, Annecy, France, and a Researcher with the LISTIC Laboratory of Computer Science, Systems, Information and Knowledge Processing, Annecy. His research interests include developments of methods and models for artificial intelligence, information processing, and image time series analysis.

Fatima Karbou received the Ph.D. degree in physical methods in remote sensing.

She is a Senior Scientist in remote sensing of Earth's surfaces. She spent several years at the CNRM research group on numerical weather prediction (NWP), working on assimilating passive microwave observations over land and sea ice surfaces within global NWP models. She joined the Snow Research Centre in early 2011. Her research interests include understanding the physics of satellite measurements, monitoring seasonal snow cover in the mountains, mountain lake dynamics, and characterizing some relevant indicators of avalanche activity using Sentinel-1 SAR image time series.

Emmanuel Trouvé (Senior Member, IEEE) received the Ph.D. degree in signal and image processing from Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1996.

He was with University Savoie Mont Blanc, Polytech Annecy-Chambéry, Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance, as an Associate Professor from 1998 to 2008, and then as a Professor of signal and image processing. His research interests include synthetic aperture radar (SAR) image processing, photogrammetry, and glacier monitoring by remote sensing.

Prof. Trouvé was the General Chair of MultiTemp 2015 (8th International Workshop on the Analysis of Multitemporal Remote Sensing Images).

Matthieu Gallet received the M.Sc. degree in fundamentals of remote sensing from the University of Paris, Paris, France, in 2021. He is currently working toward the Ph.D. degree with LISTIC, University Savoie Mont-Blanc, Annecy, France, and CEN, Météo France, Grenoble, France.

His research interests include statistical signal processing and machine learning for SAR images.