© © 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Title: A System for the Automatic Classification of Ice Sheet Subsurface Targets in Radar Sounder

Data

This paper appears in: IEEE Transactions on Geoscience and Remote Sensing

Date of Publication: 02 November 2014

Author(s): A.M. Ilisei, L. Bruzzone

Volume: 53, Issue 6 Page(s): 3260 – 3277

DOI: 10.1109/TGRS.2014.2372818

A System for the Automatic Classification of Ice Sheet Subsurface Targets in Radar Sounder Data

Ana-Maria Ilisei, Student Member, IEEE, and Lorenzo Bruzzone, Fellow, IEEE

Abstract

Exhaustive investigations of the ice sheet subsurface can be carried out by analyzing the information contained in the huge archives of radargrams acquired by dedicated radar sounder (RS) instruments. The analysis can be done by using properly designed automatic techniques for a quantitative, objective and reliable extraction of information from radargrams. Unfortunately, the definition and development of such automatic techniques has been only marginally addressed in the literature. In this paper, we propose a novel and efficient system for the automatic classification of ice subsurface targets present in radargrams. The core of the system is represented by the extraction of a set of features for target discrimination. The features are based on both the specific statistical properties of the RS signal and the spatial distribution of the ice subsurface targets. Such features are then provided as input to an automatic classifier based on Support Vector Machine (SVM). Experimental results obtained on two real-world datasets acquired by airborne-mounted RSs in large regions of Antarctica confirm the robustness and effectiveness of the proposed classification system.

Index Terms

Radar sounding, subsurface analysis, ice subsurface targets, cryosphere, signal processing, remote sensing.

I. INTRODUCTION

VER the past decades, the investigation of the ice sheets has been carried out by analyzing the ice subsurface in radargrams (or echograms) acquired by dedicated radar sounder (RS) instruments (also known as Ground Penetrating Radars (GPR) for glaciology) [1]. RSs are active instruments that can perform nonintrusive depth measurements of the subsurface structure of the ice sheets on very wide areas. Such measurements are possibile due to the use of relatively low-frequency electromagnetic waves (typically in the MF, HF, VHF frequency spectrum) and the nadir-looking acquisition geometry that enable the penetration of the wave several kilometers into the ice subsurface [1]. Since emitted, the wave suffers geometric losses with depth and attenuation caused by the dissipative properties of the subsurface [2]. The dielectric discontinuities present within the ice subsurface structure form interfaces that reflect the impinging wave. The amount of power returned to and measured by the radar receiver is recorded in radargrams.

RSs are usually operated on airborne or satellite platforms. RSs mounted on airborne platforms (e.g., MultiCoherent Radar Depth Sounder (MCoRDS) [3], Polaris [4]) are employed for sounding the Earth ice sheets and glaciers. Orbital RSs have been used only for the exploration of other planetary bodies, e.g., Moon (Lunar Radar Sounder (LRS) [5]), Mars (Shallow Radar (SHARAD) [6]). During such missions the ice sheets have been sounded on very wide areas generating a huge amount

of RS data that have been stored in archives and only partially analyzed. Moreover, the volume and heterogeneity of data is expected to grow from the increase in both terrestrial and forthcoming space-based missions.

The main approach used to the analysis of radargrams is based on visual interpretation. This approach is very time consuming, therefore difficult to apply to a large amount of data. Moreover, it is intrinsically subjective, thus leading to photointerpretation inconsistencies which make impossible a standardization of the output. Furthermore, the traditional manual approach can only be applied to an individual radargram at a time and makes it difficult a joint processing of different radargrams. This limits the possibility of analyzing radargrams that are geographically related or to fuse together different types of remotely sensed data (e.g., radargrams, Synthetic Aperture Radar (SAR) data, Digital Elevation Models (DEMs)). Such joint analysis could provide further information about the ice subsurface structure (e.g., 3D subsurface models) that is difficult to obtain with a manual approach. A solution to the limitations of the manual approach is the development of efficient automatic techniques for the analysis of radargrams. Such techniques can provide fast, quantitative and objective results. Therefore, they can be particularly important for an accurate investigation of the ice subsurface at large scale. Despite these advantages, the development of such techniques is still in an early stage, as pointed out in the review of the state of the art presented in Sec. II.

In this paper we present an advanced and effective system for the automatic classification of the whole backscattering area of the ice subsurface targets. In brief, from a physical point of view, the ice subsurface is composed by layers of ice and the underlying bedrock [1]. In radargrams, they appear as different patterns that can be recognized by their structure, continuity, depth location and reflected wave amplitude and phase. Another pattern present in radargrams corresponds to the measurements of pure noise. Therefore, *layers, bedrock* and *noise* are the targets that we aim to automatically classify (a detailed description of the ice sheet subsurface targets is given in Sec. III). Identifying these ice subsurface targets represents a first fundamental step for a subsequent more complete understanding of the ice sheets, e.g., the computation of the ice thickness, the study of archeological changes (see Sec. III for details). However, automatic approaches to the identification of the ice subsurface targets have not been sufficiently addressed by the scientific community.

The automatic classification system that we present in this paper combines advanced image processing and machine learning techniques with the knowledge about the physical distribution of the targets and fundamentals on radar wave backscattering. After an initial elevation correction step applied to the radargrams for removing the effect of fluctuating aircraft altitude, they are given as input to the system, which is made up of two main components: i) feature extraction, and ii) automatic classification based on Support Vector Machine (SVM). The feature extraction is the core of the system and also the main novel contribution of this work. The objective at this stage is to extract from the radargrams effective parameters for target discrimination (in this paper we call such parameters "features", in accordance with the pattern recognition literature; it is worth to note that a conventional term used in the glaciological and ice radar communities for "ice subsurface targets" is "ice subsurface features". However, in this paper we use the term "ice subsurface targets" to avoid possible confusion caused by the same word "feature" associated with two different fundamental concepts). We propose a set of features that are able to model and correlate the backscattering properties of the radar signal with the spatial properties of the subsurface targets. The extraction of such features is done after a detailed study of the statistical properties of the radar signal and of the spatial distribution of the ice subsurface targets. The second component of the system uses the extracted features

to perform the automatic classification of ice subsurface targets by using the SVM classifier. The main advantages of the system are: i) robustness and/or adaptiveness to the heterogeneity of radargrams as a consequence of both the features used and the learning approach employed; ii) capability to obtain objective quantitative results (i.e., exactly the same criteria are used for all radargrams, thus enabling the extraction of targets in a consistent and comparable way on all radargrams); and iii) computational speed and efficiency due to the possibility of parallelizing the sub-algorithms it is composed of. For these reasons, the system is suitable for the analysis of the ice subsurface at large scale from radargrams acquired by RS sensors with different characteristics (e.g., central frequency, bandwidth), as it will be shown in the paper. The system has been validated on two real-world datasets [3]: i) a dataset made up of 8 radargrams acquired in sequence off ≈ 400 line-km in Central Antarctica by the airborne-mounted MCoRDS instrument [7], and ii) a dataset made up of 14 radargrams acquired in parallel- and cross-track configurations over an area of about 1000 km² in SE Antarctica (Byrd Glacier) by MCoRDS2 [8]. The rest of the paper is organized as follows. Sec. II reviewes the literature related to the automatic analysis of radargrams. A complete description of the ice sheet subsurface targets is given in Sec. III. Sec. IV and Sec. V present the main components of the automatic classification system, i.e., the feature extraction techniques and the classification method based on SVM. Experimental results obtained on the two RS datasets acquired in Antarctica are reported in Sec. VI. Finally, Sec. VII discusses the capabilities and limitations of the system and proposes future developments of this work.

II. RS DATA PROCESSING AND RELATED WORK IN GLACIOLOGY

A radargram is a 2D matrix with nS rows (samples in the range/depth direction) and nT columns (traces, frames in the along-track/azimuth direction), which contains the measurements of the RS [1].

The analysis of radargrams is very challenging, as the amplitude A(i, j), $\forall i = [1..nS]$, $\forall j = [1..nT]$, may contain both useful information and noise contributions (e.g., thermal noise, speckle, clutter, multiple reflections and sidelobes artifacts) [1]. These noisy contributions may partially or even completely mask the useful signal, thus leading to wrong interpretations of the information contained in radargrams. Despite the advantages that the automatic techniques could provide (see also Sec. I), the related literature in the analysis of RS data reports only few works. An attempt to automatically estimate the polar ice thickness from airborne data is presented in [9]. Here, the authors propose two techniques for the automatic detection of the ice surface and the bedrock interfaces, i.e., i) edge-based, and ii) active contour. In [10], the technique for tracing the depth of the Holocene in Greenland is presented. The technique is semi-automated and uses image processing concepts based on histogram analysis and surface fitting to identify the transition region between the Holocene and Glacial ice. Recently, a subsurface model-based technique for the detection of ice sheet target interfaces has been presented in [11]. Besides these works, which deal with the segmentation of the ice sheet subsurface into different regions, there are a few works focused on understanding the ice stratigraphy, which is useful for ice flow modeling [12] or the isochronous characterization of the ice [13]. These objectives have motivated many efforts for developing automated methods (e.g., [14], [15], [16]). Other methods focus on the detection of water at the ice/bedrock interface. In [17], the authors treat the detection of the water as a binary classification problem which they solve by using a combination of eleven learning algorithms. Other related works regard mainly the analysis of the targets visible in radargrams by comparison with ground truth data collected during drilling campaigns or using other sensors. As an example, in [18] the authors investigate how the structure of the ice subsurface affects the wave propagation and its impact on the radargram acquisition process. They compare RS data, ice core line-scan images (which display the stratigraphy of high-scattering zones for light), crystal orientation fabrics and dielectric properties of subsurface samples collected from an ice core for assessing the type of investigated targets in radargrams. This analysis is extremely useful, as its scientific output could act as reference knowledge that along with the radargrams allow for further automatic processing.

Regarding the analysis of planetary RS data, works like [19] and [20] present techniques for the detection of layers in the shallow subsurface of the North Pole of Mars. In [21], the authors firstly propose the use of several theoretical models to characterize the statistical properties of the RS signals. Then, in order to isolate the basal returns from other echoes, they propose an algorithm based on a region growing technique that combines the results of the statistical analysis with the knowledge of the geometrical properties of the subsurface targets.

It is worth noting that the above-mentioned techniques only focus on either the detection of linear features in the ice stratigraphy or the detection of the ice/bedrock interface or the identification of the basal scattering area. Moreover, they are not designed for addressing the problem of the heterogeneity of radargrams. The existing RS datasets are often made up of radargrams characterized by different attributes (e.g., resolution) as they are typically acquired during several airborne campaigns with different sensors or with the same sensor operated at different modes (e.g., bandwidth). In this context, it is important to develop automatic systems that can accurately identify the above-mentioned targets all together and can be used in a flexible way on different types of radargrams.

III. PROPOSED SYSTEM: DEFINITION OF THE TARGET CLASSES IN RS DATA AND GENERAL ARCHITECTURE

The classification of target backscattering behavior in RS data requires a very good understanding of the structural properties of the ice subsurface and radar wave propagation [2]. As briefly introduced in the previous section, studies of the ice sheets (e.g., [1], [22]) reveal that the ice column is made up of a sequence of ice layers, characterized by different dielectric properties. They have been generated over millennia by snow accumulation (on the underlying bedrock) alternated by depositions of impurities from volcanic explosions [23], and ice flow dynamics [24], therefore have an isochronous character [25]. In radargrams they appear as spatially coherent surfaces that generate quasilinear patterns. The brightness of such patterns (which is related to the amplitude of the received wave) decreases with depth due to the attenuation through the subsurface [2]. Another physical component of the ice sheet subsurface, which is located below the layers, is the bedrock. In radargrams, the bedrock can be identified as the deepest scattering area. Note that, contrary to the real scenario in which the bedrock interface is expected to be contiguous, the bedrock scattering area visible in radargrams can be composed of disjunct regions, i.e., on some traces of the radargram the bedrock returns can be completely absent. The discontinuities are likely to be due to the loss of transmitted power through the ice column, to the wave total reflection caused by the supraglacial or englacial water [26], or to acquisition issues (e.g., clutter returns that can completely mask the return from nadir [1]). Therefore, the quality of the bedrock scattering

5

area mainly depends on the type of material, the topography, the conditions at the basal interface and the processing applied to the radar data. The bedrock completely attenuates the transmitted wave. This implies that at a depth larger than that of the bedrock the radar receiver measures only noise. This is visible in the bottom part of the radargrams as a homogeneous region characterized by the absence of relevant reflections. Another noise-like pattern, visible for a few hundred of meters above the bedrock is called echo-free zone (EFZ). The EFZ has been firstly identified and studied in [27]. Then, papers like [22], [28], [18] have provided deeper insight and recently some authors have confirmed the presence of the EFZ [29]. According to such studies, the EFZ is often seen away from the ice domes and ice divides, but in extensive areas of the ice sheets. Note that the EFZ is not an ice subsurface physical region (like the layers or bedrock), but rather a consequence of the radar acquisition process. In [18], it is suggested that at the EFZ corresponding depth the disturbances introduced by the ice flow caused an increase of the layer roughness. Such large scale roughness reduces the coherency of the reflecting surfaces, thus generating the echo-free zone (EFZ). Besides the layers, bedrock and noise regions, which are typically shown in radargrams (see Fig. 1(a)), near-bed reflectors have been recently identified as freeze-on ice [29]. These reflectors are found primarily along the high ridges at the valley heads and along the steep valley walls surrounding subglacial mountain peaks. Furthermore, when present, deep specular and strong reflections are associated with subglacial lakes [30]. Therefore, the freeze-on ice and the liquid water constitute other two subsurface targets (see Fig. 1(a)), which should be considered in the modeling of the subsurface. However, since by visual interpretation it is difficult to assess with high accuracy the freeze-on regions (if present) and given that in the datasets at hand no liquid water returns could be identified, in the following we do not consider such regions as target classes for automatic classification.

The importance of identifying the ice sheet subsurface targets has been often highlighted in the literature. In particular, the results obtained from the presented algorithm can be used for instance in studies that can further focus on the interpretation of the detected layered area only, by applying other techniques for the identification of individual layers. Regarding the EFZ, in [18] the authors state that identifying the EFZ onset is fundamental since it indicates changing archeology that shall be accounted for in the modeling of ice sheet dynamics (similar to the identification of the depth of the Holocene in Greenland [10]). Also, one can analyze the shape of the EFZ for a better understanding of its formation. Finally, the identification of the whole bedrock backscattering area can be used in geological studies for assessing the type of material the bedrock it is made of [31] or to understand the reasons for which the bedrock (i.e., the basal interface) helps estimating the topography of the bedrock, computing the thickness of the ice column and inferring information about the basal boundary conditions and processes (e.g., presence of melted ice at the interface). The last return of the bedrock marks the depth below which the losses through the subsurface (ice and bedrock) have completely attenuated the transmitted power. Thus, it can be used to derive the absorption properties of the bedrock.

In this paper we aim to develop a system for the automatic classification of *layers*, *bedrock* and *noise* (which includes also the EFZ region). An example of backscattering from these classes is given in Fig. 1(b). In order to perform the classification, the radargrams are initially altitude corrected for removing the effect of the aircraft fluctuations. Then they are given in input to the classification system, which consists of 2 main components, i.e., i) feature extraction for target description, and ii) automatic



Fig. 1. (a) Qualitative representation of ice sheet target classes typically visible in radargrams. In the presented work, only the target classes highlighted in bold, i.e., layers, bedrock, noise, are considered, whereas those highlighted in italics, i.e., water, freeze-on ice, are intentionally omitted. For details see Sec. III. (b) Example of backscattering from the *layers, bedrock* and *noise* target classes.

classification based on SVM. As it will be explained later, the system requires a mimimum amount of human interaction in the training phase, in which the values of the few system parameters should be tuned to both the characteristics of the data and the scale of the subsurface targets. However, this not a critical problem, since such parameters are directly related to properties of the targets and can be easily derived. On the other hand, after the training, the system is completely automatic. Moreover, it is important to mention the flexibility and learning capabilities of the system, e.g., depending on the radar frequency and resolution of the radargrams, different target classes with associated patterns can be identified (e.g., high resolution data allows the identification of crevasses). Therefore, one first needs to set the number of classes, appropriately model the properties of the classes in the feature extraction phase and then train the classifier to automatically recognize such classes.

IV. PROPOSED SYSTEM: FEATURE EXTRACTION

The possibility to measure similar values of reflected power from different targets (e.g., returns from deep layers and bedrock can have the same power, see Fig. 1(b)) and the noisy character of the radar images make the extraction of significant features for automatic classification a very challenging task. Here we address this problem by presenting a set of features that we chose after a detailed analysis of the amplitude fluctuation of the radar signal and of the spatial distribution of the investigated targets. For ensuring a logical flow, we structure this section in three parts. First, in Sec. IV-A we present the preliminary analysis that we performed on the statistical properties of the radar signal. Then, in Sec. IV-B we analyze the spatial distribution of the subsurface targets. Finally, in Sec. IV-C we describe in detail the procedure for extracting features that model both the statistical and the spatial properties of the radar signal and of the subsurface targets.

A. Analysis of the statistical properties of the radar signal

Similarly to [21], we first performed a *statistical analysis* of the distribution of the radar signal. We analyzed the distribution of the radar signal by empirically fitting several probability density functions (pdf), i.e., Rayleigh (R_{pdf}), Nakagami (N_{pdf}), K (K_{pdf}), Gamma (G_{pdf}), to the histogram of samples drawn from regions corresponding to the investigated target classes. The

abovementioned pdfs are parametric models, i.e., they can be described by using a finite number of parameters $\theta = (\theta_1, \theta_2, ...)$. Tab. I reports the parameters describing each of these theoretical distributions. The choice of these pdfs is motivated by their expected capability to model the amplitude fluctuations of the radar signal backscattered by different targets and/or processed with different algorithms, as it has been proven in works like [32], [21]. As such, the Rayleigh pdf (R_{pdf}) typically models the amplitude oscillation of a zero-mean additive Gaussian noise (AWGN) (e.g., this is the case of thermal noise measured by the radar in the regions with no backscattering). The Nakagami pdf (N_{pdf}) generally models amplitude radar data that have been priorly subjected to multilooking processing (for speckle reduction). The K pdf (K_{pdf}) generally guarantees good performances for fitting data from regions with bunched scatterers (e.g., this is the case of *layers* and *bedrock* returns). The Gamma pdf (G_{pdf}) is generally employed in the intensity domain (I), for fitting data whose distribution in the amplitude domain (A) follows a Nakagami pdf (we remind that $I \propto A^2$). Moreover, due to its flexibility, the Gamma pdf is likely to model data whose original distribution has been altered by possible processing. The analytical formulation of these pdfs along with the procedure for estimating their parameters are reported in Appendix A (for further details refer to [21], [32], [33]). Among all the investigated distributions, the best fitting model for each class can be chosen as the one that minimizes the

$$KL(H,M) = \sum_{A_i} H(A_i) \log \frac{H(A_i)}{M(A_i)},\tag{1}$$

where H is the real histogram of the amplitude samples and M is one of the investigated theoretical models, i.e., $M = \{R_{pdf}, N_{pdf}, K_{pdf}, G_{pdf}\}$.

Kullback-Leibler (KL) distance [34] between two distributions H and M, defined according to:

Thus, given a specific RS instrument and the related data, we can select the distribution that best fits the target classes as the model that empirically minimizes (1).

Distribution	Parameters	Parameter name	
R _{pdf}	$\theta_R = \mu_{A^2}$	mean power	
N_{pdf}	$\theta_N = (\mu_{A^2}, \beta_N)$	mean power, shape	
K_{pdf}	$\theta_K = (\mu_{A^2}, \beta_K)$	mean power, shape	
G_{pdf}	$\theta_G = (\alpha_G, \beta_G)$	scale, shape	

TABLE I THEORETICAL MODELS AND THEIR PARAMETERS

B. Analysis of the properties of the subsurface targets

In order to properly design the proposed system, we also performed a qualitative analysis of the ice subsurface representation in radargrams (see Fig. 1). This allows obtaining an approximate knowledge of the *location and spatial distribution* of the target classes, which can be then used in the feature extraction for classification. From this analysis we derived that:

- *i)* The *expected order in the range direction* of the ice sheet subsurface target classes visible in radargrams is: *layers, noise* (*EFZ*, if present), *bedrock* and *noise*. This statement has general validity, as it could be derived from Sec. III.
- *ii*) The ice subsurface targets visible in radargrams are mostly *extended in the along-track direction*, due to the isochronous character of the ice stratigraphy and the continuous shape of the bedrock.

iii) The ice subsurface targets shown in radargrams present *significant variation of backscattering along the range direction*. These variations are caused by reflections from the layered structure of the ice column and the basal interface.

Before going further, it is important to notice that the radargrams can be partitioned in two main regions, the subsurface region R_{ss} that contains all the target classes of interest (i.e., layers, bedrock, noise) and a region R_{noise} at the bottom of the radargrams that contains exclusively noise measurements. In order to speed-up the processing, we will focus the following analysis on the R_{ss} region.

C. Features that model the statistical properties of the radar signal and the geometrical distribution of the subsurface targets

The analysis on the statistical distribution of the amplitude radar data and the location and spatial distribution of ice subsurface targets that has been carried out previously enabled the identification of the features for classification described below.

- 1) Parameters of the best fitting model. Once the best fitting model for the statistical characterization of the radar signal has been identified [see Appendix A and (1)], we use the values of its parameters θ_{best} as features. In other words, if the best model is the K_{pdf} , then $\theta_{best} = \theta_K = (\mu_{A^2}, \beta_K)$, or if the best model is G_{pdf} , then $\theta_{best} = \theta_G = (\alpha_G, \beta_G)$, and so on. In order to cover the whole radargram space, for computing these features we employ a sliding window approach, as in the following. We use a rectangular window inside which we estimate the values of the desired parameters [by using the appropriate eq. among (9), (11), (13), (15)]. The window is moved over the R_{ss} region with a step of one pixel both in the along-track and range directions. In order to filter out some noisy contributions, the final value of these features at each step of the sliding window is computed by averaging the estimated values on overlapping windows. It is worth mentioning that, from a statistical and image processing point of view, the size of the rectangular window $[W_r \times W_a]$ $(range \times along-track)$ should be sufficiently small for avoiding filtering the information at the borders of the scattering classes, while the number of samples inside the rectangular window should be sufficiently large for a good estimation of the parameters of the distributions. The resolution of the radargram, the spatial distribution and possibly the knowledge of the scale of the subsurface targets should also be considered when choosing the size of the sliding window. From the qualitative analysis performed previously regarding the spatial distribution of the subsurface targets, i.e., they are elongated in the alogn-track direction and present higher backscattering variation in the range direction, we can derive that an initial constraint on the choice of the sliding window is $W_a > W_r$. This constraint allows for a more consistent averaging when applying the sliding window approach, as it ensures a high level of affinity among the samples within the window. Such observation on the choice of the values for W_r and W_a hold for all the features computed on a sliding window basis.
- 2) Texture. As pointed out in Sec. III, a qualitative analysis of the radargrams indicates that different target classes present distinctive patterns. We convert such qualitative information into a quantitative measure, by computing a texture feature. Among the many texture measures used in radar image processing (e.g., [35]), we consider the entropy Ent, which is a simple but informative measure. The entropy is a statistical measure of the uncertainty of a random variable, i.e., the more uncertain a random variable, the higher its entropy value. Accordingly, when computing the entropy of the samples of the radargram that belong to the *layers* and *bedrock* classes, it is expected to obtain a high value. This is due to the fact that the amplitudes of the backscattered waves in these regions can have very large dynamic range since they also

depend on the structure and dielectric properties of the investigated targets which can vary significantly within the ice subsurface. On the other hand, the *noise* regions are characterized by relatively similar values (resulting in a less textured portion in the radargram), therefore their entropy is relatively lower. Thus, *Ent* helps in differentiating the subsurface targets on the basis of specific patterns that are characterized by the probability of appearance $p(C_i)$ computed in a local neighborhood $W_r \times W_a$, according to the sliding window approach described previously, i.e.,

$$Ent = -\sum_{C_i \in W_r \times W_a} p(C_i) \log_2 p(C_i),$$
⁽²⁾

where C is a quantized version of the log-amplitude radargram $C = Q(10 * \log_{10}(A))$ and $Q(\cdot)$ is the uniform quantization operation on N_q levels. Note that the quantization is a common operation used to reduce the very large dynamic range of the radar data to only N_q distinct values (e.g., [36]).

- 3) Kullback-Leibler distance between the distribution of the samples of the target classes and of the noise. Using the same sliding window approach and the output of the statistical analysis, we generate a feature that statistically models the distance between the measured bakscattering and the background noise. The literature suggests that a potential such statistical distance measure applicable to RS data is the KL distance [21]. Accordingly, we computed the KL distance of the radargram $KL_{R_{ss}}$, by applying (1) to the R_{ss} region. Here, H is the histogram of the amplitude samples within the sliding window and M is the noise model fitted on the samples of the window. The values of the noise parameters have been estimated (as explained in Appendix A) on the R_{noise} region. Note that the $KL_{R_{ss}}$ is a smoothed version of the input radargram, in which the most scatterable subsurface target are highlighted. Therefore, $KL_{R_{ss}}$ represents a good measure to discriminate between the samples belonging to high backscattering areas (i.e., *layers* and *bedrock*) and those of noise.
- 4) Range position of the subsurface targets. Intuitively, useful information that could help in discriminating the different types of backscattering classes is the distance P_r of the subsurface targets with respect to the air/ice interface $surf_a$, see (3). $surf_a$ is detected automatically for each trace of the radargram as the position of the maximum return along the trace. This is a fast and simple approach that has been employed in other works for the analysis of airborne acquisitions (e.g., [16]).

$$P_r(i,j) = [i - surf_a(j)], \forall i > surf_a, \forall j.$$
(3)

5) *Relational feature*. A less intuitive feature relates the position of the samples in the range direction with their backscattering strength. For this reason we call it relational feature R_{down} . Its objective is mainly to enable the separation between the returns of the classes with high backscattering, i.e., *layers* and *bedrock*. To this aim, a first requirement is to isolate in the radargram these high backscattering classes. This is achieved by exploiting the property of the statistical $KL_{R_{ss}}$ distance



Fig. 2. Qualitative example of trace j_{eg} for the KL_{Rss} and KL_{bin} measures (top plot), and corresponding relational feature R_{down} (bottom plot).

measure, i.e., the fact that it highlights the most scatterable targets. In particular, we threshold the $KL_{R_{ss}}$ as follows:

$$\mathrm{KL}_{bin}(i,j) = \begin{cases} 1 & \text{if } \mathrm{KL}_{R_{ss}}(i,j) \ge thr_{\mathrm{KL}} \cdot \mu_{\mathrm{KL}_{R_{noise}}}, \\ 0 & \text{otherwise}, \end{cases}$$
(4)

 $\forall i, j,$

where KL_{bin} is the resultant thresholded (binary) KL_{Rss} measure, $\mu_{\text{KL}_{Rnoise}}$ is the mean of the samples of the KL_{Rnoise} [KL_{Rnoise} has been generated by applying (1) to R_{noise}] and thr_{KL} is a user defined threshold that tunes the degree of similarity between the samples of the KL_{Rnoise} and those of the KL_{Rss} measures. Note that, since the range of possible values of the threshold is $thr_{KL} > 0$, choosing an optimal value for the threshold requires a minimum amount of human interaction. In order to filter out only the regions of the KL_{Rss} corresponding to the class *noise* in the amplitude domain, a low value of the thr_{KL} is preferable. Otherwise, by choosing a too large value, the risk is to filter also high backscattering contributions. After the thresholding operation, the discrimination between samples belonging to different backscattering classes is achieved by taking into account their expected order in the range direction (see Sec. IV-B). In particular, R_{down} is generated in a columnwise manner, starting from $surf_a$ (with the initial condition $s_{down}[surf_a(j), j] = 1, \forall j$), and computing a constrained cumulative sum while moving downwards over the KL_{bin} map. The constraint is to sum 0 instead of 1 at the positions where KL_{bin} = 1. Qualitatively, by looking downwards in the range direction, each trace of the R_{down} feature has monotonically increasing values, with a behavior depending on the measured backscattering contribution (see Fig. 2, which represents the vertical profile of a generic trace j_{eq} of the R_{down} feature).

In the presented system, all the above-mentioned features are given as input to the classification algorithm. These features are the amplitude of the backscattering A, the parameters of the best fitting model θ_{best} , the entropy Ent, the Kullback-Leibler distance $KL_{R_{ss}}$, the range position of the subsurface targets P_r and the relational feature R_{down} . Therefore, the resulting feature vector \mathbf{x} can be defined as:

$$\mathbf{x} = [A, \theta_{best}, \mathrm{KL}_{R_{ss}}, Ent, P_r, R_{down}].$$
(5)

V. PROPOSED SYSTEM: AUTOMATIC CLASSIFICATION WITH SUPPORT VECTOR MACHINE

The feature vector is given as input to a supervised automatic classifier. Based on a set of labeled training samples, the aim of the automatic classifier is to build a model (characterized by a set of parameters) which can accurately predict the labels for unknown (test) target samples. There are several automatic classifiers presented in the literature among which we chose the SVM [37]. The SVM is currently the state of the art in the automatic classification of remotely sensed data [38]. Our choice is also due to the fact that the SVM has many properties very useful for solving our classification problem. Among these properties, we mention: i) very good generalization capability (it is able to avoid overfitting the model on the training samples); ii) capability to solve non-linearly separable problems in the original feature space; and iii) sparseness and uniqueness of the solution of the learning problem.

The SVM is a binary classifier. However, multiclass problems can also be solved by employing architectures made up of different binary SVMs (e.g., one against one (OAO), one against all (OAA)) [39], [40]. Here, the basic principle of the binary SVM is only briefly summarized. For solving non-linearly separable problems in the original feature space, the SVM uses a mapping function to project the samples into a higher dimensional feature space in which they are separable by hyperplanes. The mapping is done implicitly by a kernel function (e.g., linear, polynomial, gaussian) and the classification is performed after optimizing a convex objective function during the training phase of the SVM. The convexity of the objective function guarantees a unique solution, which is the optimal decision boundary between classes. Such decision boundary is the hyperplane in the transformed kernel space that maximizes the geometric margin between the training samples of the two classes taking into account a regularization term. There are several studies that treat both theoretical and practical aspects related to the use of the SVM (e.g., [38], [41]). As this kind of analysis is out of the scope of this paper, we here provide only the analytical formulation of the objective function to be optimized in the learning process of the SVM and the corresponding decision boundary that have been used by the presented system. The dual formulation used for solving the constrained optimization problem associated with the training of the SVM is given by:

$$\begin{cases} \max_{\alpha} \sum_{i=1}^{N_s} \alpha_i - \frac{1}{2} \sum_{i=1}^{N_s} \sum_{i=1}^{N_s} y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{subject to: } \sum_{i=1}^{N_s} y_i \alpha_i = 0, \quad 0 \le \alpha_i \le C, \quad 1 \le i \le N_s, \end{cases}$$
(6)

where N_s is the number of training samples characterized by the pairs (\mathbf{x}_i, y_i) . \mathbf{x}_i is the feature vector (see (5)) and y_i is the label associated to the sample *i*. α_i are the Lagrange multipliers involved in the optimization process and *C*, also called error penalization term, represents the cost associated to a wrong classification. *C* and the parameters of the kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$ constitute the set of SVM model parameters that have to be optimized during the learning process. After the optimization, the final decision boundary (solution) of the SVM is given by the following equation:

$$f(\mathbf{x}) = \sum_{i \in SV} y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b, \tag{7}$$

where b is the bias term, which measures the distance of the hyperplane from the origin. Note that the sparseness of the solution is explained by the fact that only a subset of samples $i \in SV$ associated to non-zero Lagrange multipliers, i.e., the



Fig. 3. Location and flight lines for the acquisition of the MCoRDS and MCoRDS2 radargrams investigated in this paper.

support vectors, are necessary in the definition of the separation hyperplane.

VI. EXPERIMENTAL RESULTS

We applied the presented algorithm to two datasets acquired by the MultiCoherent Radar Depth Sounder instrument (MCoRDS, owned by the Center of Remote Sensing of Ice Sheets (CReSIS) [3]). The datasets where acquired by the instrument operated with different bandwidths, i.e., Bw = 9.5 MHz and Bw = 30 MHz, in different regions of Antarctica. In order to distinguish the data, when operated with Bw = 9.5 MHz, the instrument and the dataset are called MCoRDS, while when operated with Bw = 30 MHz, they are called MCoRDS2. The approximate positions and the paths followed by the aircrafts carrying the instruments, MCoRDS and MCoRDS2, are shown in Fig. 3, in green and red, respectively. In the following we present: i) the description of the two datasets, ii) the results of the statistical analysis of the radar signal, iii) the experimental setup employed in the training phase of the SVM classifier, iv) the classification results and v) the computational efficiency obtained by applying the presented technique to the two datasets.

A. Dataset description

The first considered dataset (MCoRDS) was acquired during the sounding campaign conducted in Central Antarctica in November 2010 [7]. It is made up of 8 radargrams acquired in sequence, i.e., from (-86.00°N, -15.67°E) to (-86.02°N, 29.45°E), over a distance of \approx 400 line-km (which corresponds to nT = 27350 traces). The instrument was flown on a jet aircraft (DC-8) at high altitude (H \approx 7000 m). The central frequency of the instrument and the bandwidth are $f_c = 193.5$ MHz and Bw = 9.5 MHz, respectively.

The second dataset (MCoRDS2) was acquired at $f_c = 193.5$ MHz with Bw = 30 MHz. The instrument was flown on a TO aircraft at a relatively low altitude (H ≈ 500 m) in parallel and cross-track configurations over an area of around 1000 km², i.e., within (-80.93°N, 145.72°E) and (-80.40°N, 148.10°E), over the Byrd Glacier in Antarctica, in December 2011 [8]. For obtaining best quality dynamic range, the MCoRDS2 dataset has been generated by multiplexing in time two types of data: i) signals collected from the shallow subsurface, acquired by using a waveform playlist (wpl) coupled with low gain channel (LGC) and a pulse duration $T_{ps} = 1\mu s$, and ii) signals collected from the deep subsurface, acquired by using a wpl coupled with high gain channel (HGC) and pulse duration $T_{pd} = 10\mu s$ [42], [43]. However, this combination introduces a certain

Parameter	MCoRDS	MCoRDS2
Location	Central Antarctica	Byrd Glacier Antarctica
Platform and altitude (H)	DC-8 @ H \approx 7000 m	TO @ $H \approx 500 \text{ m}$
Number of radargrams	8	14
Acquisition strategy	sequence	parallel and cross-track
Distance / Area	400 line-km	1000 km ²
Central frequency (f_c)	193.9 MHz	193.9 MHz
Bandwidth (Bw)	9.5 MHz	30 MHz
Transmitted power	550 W	1050 W
Acquisition	14 bit ADC @ 111 MHz	14 bit ADC @ 111 MHz
Dynamic range	wpl	wpl with LGC and HGC
Range resolution in ice (R_r)	13.6 m	4.3 m
Along-track resolution (R_a)	25 m	25 m
Total number of samples (nS)	410	1200
Total number of traces (nT)	27350	17093

 TABLE II

 PARAMETERS AND CHARACTERISTICS OF THE INVESTIGATED DATASETS

amount of heterogeneity between the radiometric quality of the data acquired in shallow and deep modes, which is difficult to handle at data processing level. Considering this and the fact that the low gain channel acquires data exclusively from the first km within the subsurface (i.e., class *layers*), and given that our purpose is the classification of ice subsurface targets, in our analysis we investigated data acquired only with the high gain channel, which contains returns belonging to all target classes, i.e., *layers*, *bedrock*, *noise*.

Regarding the quality of data, several preprocessing techniques have been applied in order to obtain improved resolution. In particular, pulse compression and windowing algorithms (e.g., 20% Tuckey window in the time domain, with widening factor $k_t = 1.53$) have been used to improve the range resolution while suppressing the sidelobe level. SAR processing has been applied to improve the along-track resolution and for clutter removal, and multilooking processing (11 looks in the along-track direction and 1 look in the range direction) for despeckling. Also, a minimum variance distortionless response (MVDR) [44] algorithm has been applied to data to suppress clutter contributions coming from the cross-track direction. It is noting that all these processing techniques affect the statistical models to be used for modeling the fluctuation of the investigated amplitude radar signal (see Sec. IV-A and Appendix A).

The parameters of the acquisition systems and the main characteristics of the data are reported in Tab. II. Fig. 4 shows the subsurface region R_{ss} of the investigated datasets.

B. Results of the statistical analysis of the radar signal

In the following, the results of the statistical analysis performed by fitting the Rayleigh, Nakagami, K and Gamma pdfs to the amplitude radar signal are presented. Fig. 5 shows the regions that have been selected for the analysis from each target class, from a portion of the (a) MCoRDS and (b) MCoRDS2 datasets. Note that in the figures the values are reported in dB (for visibility), while the statistical analysis has been performed on the normalized amplitude data. Also, note that we considered the *EFZ* class individually (i.e., not merged with the *noise* class). This class has been intentionally selected separately, since another objective in these experiments is to verify also from a statistical analysis are sufficiently representative, in the fitting process for both datasets we picked a very large number of samples per class (see Tab. III). The fitting performances, which have been evaluated in terms of Kullback-Leibler distance, [34], are reported in Tab. IV, in which the best fitting results for



Fig. 4. Datasets investigated: a) MCoRDS $[nS = 410 \times nT = 27350]$, b) MCoRDS2 $[nS = 1200 \times nT = 17093]$. Values are in dB. The figures are stretched (color adjusted) and vertically exagerated in order to highlight the regions of interest of the subsurface. The upper black region corresponds to the free space above the surface return $surf_a$ (for the MCoRDS2 dataset $surf_a$ is computed using the radargram acquired by the low-gain channel data). The white band in the shallow subsurface (first 285 samples \simeq 798 m below $surf_a$) of the MCoRDS2 dataset corresponds to the data acquired by the low-gain channel, which is not investigated in our analysis.



Fig. 5. Samples picked manually from each target class on a portion of radargram from (a) the MCoRDS dataset and (b) the MCoRDS2 dataset. In the figures, each color corresponds to a different target class, i.e., red - *layers*, green - *EFZ*, blue - *bedrock*, yellow - *noise*.

 TABLE III

 Number of picked samples per class (corresponding to the regions highlighted in Fig. 5) used in the statistical analysis.

	Number of picked samples				
Target class	MCoRDS MCoRDS2				
layers	38351	43979			
EFZ	12257	16314			
bedrock	9321	22710			
noise	21381	32754			

TABLE IV

FITTING PERFORMANCES IN TERMS OF KULLBACK-LEIBLER DISTANCE (DIMENSIONLESS) OF THE RAYLEIGH, NAKAGAMI, K AND GAMMA DISTRIBUTIONS TO THE SAMPLE AMPLITUDE DATA FOR LAYERS, EFZ, BEDROCK AND NOISE CLASSES. THE BEST RESULTS (SMALLEST VALUES ON EACH COLUMN FOR EACH DATASET) ARE HIGHLIGHTED IN BOLD.

		Target class				
Dataset	Distribution	layers	EFZ	bedrock	noise	
	Rayleigh	0.7097	0.0809	0.6117	0.0815	
MCaDDS	Nakagami	0.1395	0.0138	0.3257	0.0127	
MCORDS	K	0.0212	0.0974	0.1442	0.0990	
	Gamma	0.0263	0.0025	0.1494	0.0015	
	Rayleigh	0.0840	0.1835	0.1433	0.2691	
MCoRDS2	Nakagami	0.0844	0.0158	0.1357	0.0084	
	K	0.0062	0.2095	0.0796	0.2995	
	Gamma	0.0029	0.0017	0.0578	0.0007	

each class (which have been derived as defined in Sec. IV-A) are highlighted in bold. Such results point out that in almost all the cases, the best fitting model is the Gamma pdf. The exceptions are for the classes *layers* and *bedrock* of the MCoRDS dataset, where the K pdf fits slightly better than the Gamma pdf (difference at the third decimal). However, given the overall very good performances of the Gamma pdf (see also Fig. 6) and the fast computation time in estimating its parameters (i.e., two analytical formulas, see (15), instead of the iterative approach employed for the K pdf, see (13)), in the following, the Gamma pdf is considered as the most suitable fitting model for all classes for both datasets. Note that this is in disagreement with theoretical grounds in radar signal distribution [32] and with the results obtained from applying a similar approach to other RS datasets [21]. For instance, in [32] it is analytically proven that in the regions of no backscattering, e.g., *noise*, the histogram of samples follows a Rayleigh distribution, which is confirmed on a subset of SHARAD radargrams in [21]. However, it is important to recall that our results have been obtained by applying the statistical analysis to data that have been preprocessed (for clutter and sidelobe reduction) and the preprocessing operations changed the data properties with respect to the datasets investigated in other studies. The qualitative results shown in Fig. 6 indicate that this preprocessing has changed the original Rayleigh distribution that can be better modeled by the Gamma pdf.

It is also worth to analyze the results reported qualitatively in Fig. 7, which shows (a) the summary of the fitted Gamma models to all target classes, and (b) the fitted Gamma pdfs to the *noise* and *EFZ* classes. These results refer to the MCoRDS2 dataset, but similar results (which are not reported in the paper for space constraints) have been obtained on the MCoRDS dataset. The plot in Fig. 7(a) indicates the large difference between the distributions of the *EFZ/noise* and *layers/bedrock* classes, and the very large dynamic range characterizing the radar signal. Fig. 7(b) points out the similarity of the two Gamma pdfs modeling the *EFZ* and *noise* samples. This similarity confirms also from a statistical point of view the validity of the hypothesis that in the automatic classification of ice subsurface targets, the *EFZ* and *noise* classes are merged within a single no backscattering target class, from now on called *noise*.

C. Experimental setup

From the considerations made above on the type of data and the scale of the subsurface features, the values of the system parameters selected in our experiments are: $W_a = 14$ and $W_r = 7$ samples, $N_q = 256$ levels and $thr_{KL} = 10$. According to our previous analysis, the Gamma pdf is the best fitting model for all the classes. We therefore extract as features its

16



Fig. 6. Empirical and estimated (with the maximul likelihood technique) distributions for each target class for the (left) MCoRDS and (right) MCoRDS2 datasets.



Fig. 7. Fitted Gamma distributions for (a) all target classes, (b) the EFZ and noise classes. Results are shown for the MCoRDS2 dataset.

parameters, i.e., $\theta_{best} = \theta_G = (\alpha_G, \beta_G)$, as explained in Sec. IV-C. These are shown along side the amplitude radargrams (which are converted in dB for visibility), the KL_{Rss} maps, and *Ent* in Fig. 8, for a portion of the MCoRDS dataset at left, and MCoRDS2 dataset at right.

The set of labeled samples for training and testing the SVM was created by defining a reference map of the subsurface. This was done by manually selecting, according to an accurate visual analysis of the radargrams, the regions corresponding to the various target classes. A subset of the reference samples along with the corresponding features are given in input to the SVM classifier for training (we recall that a generic sample **x** is characterized by seven features, i.e., $\mathbf{x} = [A, \alpha_G, \beta_G, \mathrm{KL}_{R_{ss}}, Ent, P_r, R_{down}]$). The subset of reference samples is chosen in order to take into account the variability of the subsurface targets in the alongtrack direction (e.g., at some locations the bedrock is deeper than in others). We split the dataset (and the reference map) in N vertical tiles, from which we collect randomly 1% of the samples belonging to each class, to be used in the learning phase. Then, the samples collected from $N_{CV} = 2N/3$ tiles are used in a training $k - fold \ cross-validation$ algorithm for selecting the SVM model parameters. In our experiments, we used a Gaussian radial basis function (RBF) kernel for the SVM. This choice is motivated by the fact that the RBF kernel is typically more flexible than the linear kernel and it usually outperforms the polynomial kernel in convergence time [41]. Therefore, the SVM model parameters are the penalty error term C and the γ parameter of the RBF kernel. C and γ are tested by performing a grid-search model selection. C is tested between $[10^{-3}..10^6]$ with a logarithmic step size, and γ is tested with 10 values in logarithmic scale, with central value $\gamma_c = 1/(2 * \sigma^2)$, where σ is the average distance between each pair of classes. Then, for testing the SVM on unknown samples, we chose the values C_T and γ_T that provided in average (on the k folds) the highest classification accuracy. The test samples are collected from the remaining $N_T = N - N_{CV}$ tiles. N is chosen depending on the number of traces nT available in the considerend dataset, i.e., N = 99 for the MCoRDS (with nT = 27350) and N = 66 for the MCoRDS2 dataset (with nT = 17093). This implies $N_{CV} = 66$ and $N_T = 33$ tiles for the MCoRDS dataset, $N_{CV} = 44$ and $N_T = 22$ tiles for the MCoRDS2 dataset, and a number of traces per tile $nT_{tile} \in [250..300]$. The number of folds is k = 11. Tab. V reports the number of samples per class used for the cross-validation and included in the test sets for the MCoRDS and MCoRDS2 datasets.



Fig. 8. Examples of extracted features. The features at the left side of the figure are (a) the radargram, (c) the shape parameter of the Gamma distribution, (e) the Entropy and (g) the $KL_{R_{ss}}$ measure on a portion of radargram (\approx 30 line-km) of the MCoRDS dataset. The features at the right side are (b) the radargram, (d) the shape parameter of the Gamma distribution, (f) the Entropy and (h) the $KL_{R_{ss}}$ measure on a portion of radargram (\approx 60 line-km) of the MCoRDS2 dataset. The radargram (\approx 60 line-km) of the MCoRDS2 dataset. The radargrams are in dB, stretched and vertically exagerated to improve visibility.

38913

96776

48698

79679

Total

 TABLE V

 NUMBER OF REFERENCE SAMPLES PER CLASS USED IN THE CROSS-VALIDATION AND TEST SETS.

D. Classification results

In the following, the analysis of the training (k-fold cross-validation) and test results is given. Tab. VI reports the average error matrix [45] computed after applying the cross-validation algorithm to k = 11 folds on the MCoRDS dataset. The average accuracy (AA), the corresponding standard deviation (STDEV), the average user accuracy (AUA) and the average producer accuracy (APA) for each class are also reported. The average accuracy values, i.e., AA, AUA and APA, are computed as the mean values of the overall accuracy (OA), user accuracy (UA) and producer accuracy (PA), respectively, calculated on each validation fold. The OA quantifies the overall goodness of the classifier. The UAs represent the percentage of samples correctly labeled in the classification map for each class. The PAs provide for a given class in the reference map, the percentage of samples correctly labeled in the classification map. Tab. VII shows the same information for the MCoRDS2 dataset. The cross-validation algorithm provides $C_T = 10^6$ and $\gamma_T = 2.08$ for the MCoRDS dataset, and $C_T = 10^3$ and $\gamma_T = 8.88$ for the MCoRDS2 dataset. The error matrices on the test sets along with the correspondent UA, PA and OA are reported in Tab. VIII and in Tab. IX, for MCoRDS and for MCoRDS2 dataset, respectively.

By analyzing the tables, one can see that the low values of the standard deviation (i.e., 0.41 for MCoRDS and 0.73 for MCoRDS2) confirm the robustness of the presented system to the random choice of the samples used in the k folds of the cross-validation algorithm. By comparing Tab. VI and Tab. VIII, one can see that, for the MCoRDS dataset, the AUA and the UA, and the APA and PA, respectively, have similar values. This means that the overall variability of the samples has been well captured in the training phase and proves that the selected SVM model for testing the system capabilities is not biased. The same observations hold for the MCoRDS2 dataset (see Tab. VII and Tab. IX). Moreover, we obtained values of OA> 97% (i.e., 99.09% for MCoRDS and 97.93% for MCoRDS2), which are very satisfactory, especially when considering the type and scale of the investigated targets, the noisy character of the data, and the fact that after training, the system is completely automatic. The effectiveness of the system is also proven by the high values of UA and PA, obtained on both datasets. From the tables one can see that the few errors are mainly due to a wrong classification of some returns at the interfaces between the classes (e.g., 193 out of 7980 bedrock samples and 70 out of 8930 layers samples are labeled as noise samples for the MCoRDS dataset, whereas for the MCoRDS2 dataset 219 and 109 noise samples out of 19808 are labeled as layers and bedrock samples, respectively). Such errors are mainly caused by the sliding window approach. Due to its intrinsic low pass filtering effect, in the layers and bedrock regions, it tends to slightly overestimate the areas with high backscattering and to underestimate the areas with low backscattering (where the signal amplitude is close to the measured background noise). These effects can be seen in the final classification maps in Fig. 9 for the MCoRDS dataset and in Fig. 10 for the MCoRDS2

dataset. For a better understanding, they are also highlighted in Fig. 11(d). On the other hand, it is important to note that in the regions characterized by deep and strongly scattering layers followed in range by the absence of bedrock returns (see an example of such radargram in Fig. 11(a)), the classifier is able to perform an accurate classification (see Fig. 11(b)). In such cases, the classifier mostly relies on the relational feature, which, by integrating both the knowledge of the radar signal statistical properties and the position in the range direction of the subsurface targets, is generally able to correctly discriminate the samples. The importance and effectiveness of the relational feature in our classification problem has been confirmed by the unsatisfactory results obtained in initial experiments in which the relational feature has been omitted from the set of extracted features used in the learning phase (i.e., $\mathbf{x} = [A, \alpha_G, \beta_G, KL_{R_{ss}}, Ent, P_r]$). In particular, we obtained a lower overall accuracy and poor quality classification maps. An example of such classification map is shown in Fig. 11(c). By comparing this map with that obtained by using all the features (i.e., $\mathbf{x} = [A, \alpha_G, \beta_G, KL_{R_{ss}}, Ent, P_r, R_{down}]$, see Fig. 11(b)), one can easily understand the effectiveness of the proposed relational feature.

TABLE VI Average (on k = 11 folds) error matrix of the samples of the cross-validation folds (MCoRDS dataset).

		Predicted samples				
		layers	bedrock	noise	Total	APA(%)
nce	layers	3417	0	8	3425	99.75
ere	bedrock	2	1645	25	1672	98.37
Ref	noise	7	8	2129	2144	99.27
	Total	3426	1653	2162	#S	=7241
	$\Delta I I \Delta (\%)$	99.75	99.54	98 54	AA=	99.28%
	<i>non</i> (<i>n</i>)	<i>))</i> .15	JJ.54	70.54	STD	EV=0.41

TABLE VII AVERAGE (ON k = 11 FOLDS) ERROR MATRIX OF THE SAMPLES OF THE CROSS-VALIDATION FOLDS (MCoRDS2 DATASET).

			Predicted samples				
			layers	bedrock	noise	Total	APA(%)
	nce es	layers	3958	6	61	4025	98.35
	ere	bedrock	7	1127	66	1200	94.95
	Ref sau	noise	51	20	3502	3573	98.01
		Total	4016	1153	3629	#S	=8798
		$\Delta I I \Delta (\%)$	98 55	97 78	96.51	AA=	97.60%
		11011 (70)	70.55	21.10	70.51	STD	EV=0.73

 TABLE VIII

 ERROR MATRIX ON THE TEST SAMPLES (MCORDS DATASET).

		Predicted samples				
		layers	bedrock	noise	Total	APA(%)
nce es	layers	18839	21	70	18930	99.51
ere	bedrock	5	7782	193	7980	97.51
Ref	noise	17	46	11940	12003	99.48
	Total	18861	7849	12203	#S=	38913
	AUA (%)	99.88	99.15	97.84	OA=	99.09%

E. Analysis of the computational load

From a computational point of view, an important characteristic of the presented system is its ability to process in a fast way large amount of data. This is due to the fact that the algorithms included in the system can be parallelized. Thus, it is possible



Fig. 9. Examples of (a) and (b) radargrams, and (c) and (d) corresponding classification maps generated with the presented algorithm (MCoRDS dataset). The radargrams are in dB, stretched and vertically exagerated to improve visibility. In the classification maps, each color represents a different target class, i.e., black - *free space*, blue - *layers*, red - *bedrock*, yellow - *noise*.



Fig. 10. Examples of (a) and (b) radargrams, and (c) and (d) corresponding classification maps generated with the presented algorithm (MCoRDS2 dataset). The radargrams are in dB, stretched and vertically exagerated to improve visibility. In the classification maps, each color represents a different target class, i.e., black - *free space*, blue - *layers*, red - *bedrock*, yellow - *noise*.



Fig. 11. Example of (a) radargram (MCoRDS dataset) showing a particular subsurface pattern, i.e., deep and strong backscattering layers and partially missing bedrock area, (b) corresponding classification map obtained by training the SVM with all the features presented in this paper ($\mathbf{x} = \{A, \alpha_G, \beta_G, KL_{R_{ss}}, Ent, P_r, R_{down}\}$), (c) corresponding classification map obtained by training the SVM with a subset of the presented features, i.e., which does not contain the relational feature R_{down} (i.e., $\mathbf{x} = \{A, \alpha_G, \beta_G, KL_{R_{ss}}, Ent, P_r\}$), (d) portions of radargram and classification map highlighting the effect of the sliding window approach (note that the low-pass filtering effect results in a slight underestimation and overestimation of the layers and bedrock classes, respectively, at their interfaces with the noise region). The radargrams are in dB, stretched and vertically exagerated to improve visibility. In the classification maps, each color represents a different target class, i.e., black - *free space*, blue - *layers*, red - *bedrock*, yellow - *noise*.



Fig. 12. Classification maps corresponding to (a) the MCoRDS dataset [radargrams in dB shown in Fig. 4(a)] and (b) the MCoRDS2 dataset [radargrams in dB shown in Fig. 4(b)]. In the classification maps, each color represents a different target class, i.e., black - *free space*, blue - *layers*, red - *bedrock*, yellow - *noise*.

		Predicted samples				
		layers	bedrock	noise	Total	APA(%)
nce es	layers	21918	15	357	22290	98.33
fere	bedrock	17	6290	293	6600	95.30
Ref sa	noise	219	109	19480	19808	98.34
	Total	22154	6414	20130	#S=	48698
	AUA (%)	<i>(b</i>) 98.93 98.07 96.77		96.77	OA=	97.93%

 TABLE IX

 Error matrix on the test samples (MCoRDS2 dataset).

to take advantage of the latest technology (e.g., clusters of large-storage and high-power computers (cpus)) in order to faster achieve the desired performances. The feature extraction and the training of the classifier can be easily split into several subtasks to be given to different cpus that can run in parallel. Given that the features are computed with a sliding window approach (which is characterized by the fact that the computations within the windows are independent), one can use a cluster of cpus to perform such computations in parallel. In the grid-search selection of the SVM model parameters with the cross-validation algorithm, both the operations within each cross-validation fold and the computations at each intersection point of the grid are independent. This enables parallelizing the algorithm also in the training phase of the classifier, which otherwise has a time complexity in the order of $O(N_s^3)$ [46]. Moreover, the feature extraction and the training of the classifier are operations that can be computed only once, in offline mode. Once the SVM model has been selected, the only online/real-time operation is the classification of new samples, which can also be performed by several cpus in parallel. In our experiments, the computational capabilities of the presented system have been proven by using a cluster of 192 cpus (@2.05 GHz) which performed all the operations per dataset in \approx 5 hours. This is a reasonable computation time if we consider the very large amount of data that has been processed. Moreover, note that the offline computations (feature extraction and SVM training with cross-validation) require about 98% of this amount of time, while the generation of the classification maps for the whole datasets (after the training phase) require only few minutes. In general, we expect to require a new training of the classifier only when either the acquisition mode or the pre-processing phase of the data are changed. Another advantage of the presented system is the fact that it can be easily tuned for analyzing different RS datasets, since it involves a small number of parameters in the overall classification algorithm (i.e., W_a , W_r , thr_{KL}).

VII. CONCLUSION

In this paper, a novel system for the automatic classification of ice sheet subsurface targets has been presented. The system relies on advanced image processing and machine learning techniques to efficiently extract the information contained in radargrams. The presented system is made up of two main components, i.e, i) feature extraction and ii) automatic classification based on SVM. The feature extraction for ice sheet subsurface target description is the main component of the system, which also represents one of the main contributions of this work. The features extracted take into account both the statistical properties of the measured radar signal and the spatial properties of the subsurface targets. Along with the original amplitude data, several features have been identified and extracted, i.e., the parameters of the best fitting model, the entropy, the Kullback-Leibler distance, the range position of the ice subsurface targets and the relational feature. The extracted features have been given as input to an automatic classifier based on SVM to obtain the final classification maps.

The main characteristics of the presented system are: i) robustness and/or adaptiveness to the heterogeneity of radargrams as a consequence of the features used and the learning approach employed; ii) capability to obtain objective and quantitative results on large amount of data; and iii) capability to process large archives of data due to the computational efficiency and the possibility of parallelizing the subalgorithms. These have been proven by applying the algorithm to two real-world datasets acquired by the MCoRDS instrument operated with different parameters (i.e., bandwidth) in different regions of Antarctica. For both datasets, covering ≈ 400 line-km, the system provided in few hours (≈ 5 hours per dataset) high quality classification maps with an overall accuracy greater than 97%. This is a very satisfactory result, considering the type and scale of the investigated targets, the noisy character of the radar data, and the fact that the algorithm is nearly completely automatic. More precisely, the system requires a minimum amount of user interaction in the training phase of the classifier, wheareas in the operational phase (classification of new data), it is completely automatic.

The output of the system can be used for estimating the extent of the subsurface targets both in the range and along-track directions (e.g., ice layered area thickness, bedrock scattering area distribution). Furthermore, when the spatial sampling allows it (e.g., sufficiently dense grid of tracks followed by the instrument), such output can be used along with appropriate RS data integration techniques (e.g., based on standard interpolation algorithms) for generating 3D models of the subsurface, useful for the estimation of the ice subsurface targets in all dimensions. This can also help to detect critical basal boundary conditions and study changing archeology or geology. Therefore, the automatic classification of the subsurface targets is an initial essential step for the further development of more elaborate analyses of the ice subsurface.

As future work, we aim to use the output classification maps, in particular at the traces (and neighborhoods) where more tracks overlap, for defining a reliable postprocessing technique for removing outliers and finally assessing a unique solution at the corresponding lat-long coordinates in the range direction. Another objective is to check the applicability of the system to radargrams acquired in Greenland. Also we aim to study the possibility to tune the presented system in order to adapt it to the detection of subsurface targets visible in radargrams acquired in other icy regions (e.g., glaciers).

VIII. ACKNOWLEDGEMENTS

We acknowledge the use of data and/or data products from CReSIS generated with support from NSF grant ANT-0424589 and NASA grant NNX10AT68G. The authors would also like to thank P. Gogineni for publishing the data [3], the Italian Space Agency (ASI) for supporting this work and the anonymous reviewers for their comments, which helped to improve this paper.

APPENDIX

In this Appendix, the theoretical distributions used in the statistical analysis of the radar signal are briefly described:

- The analytical equation of the *Rayleigh* pdf is given by:

$$R_{pdf}(A) = \frac{2A}{\mu_{A^2}} \exp\left[-\frac{A^2}{\mu_{A^2}}\right],$$
(8)

where the parameter of the distribution μ_{A^2} is the mean power of the signal and can be estimated using the Maximum Likelihood Estimation (MLE) approach as:

$$\tilde{\mu}_{A^2} = E\left\{A^2\right\},\tag{9}$$

with $E\left\{\cdot\right\}$ denoting the expectation operation.

- The analytical equation of the *Nakagami* pdf is given by:

$$N_{pdf}(A) = 2\left(\frac{\beta_N}{\mu_{A^2}}\right)^{\beta_N} \frac{A^{2\beta_N - 1}}{\Gamma(\beta_N)} \exp\left[-\frac{\beta_N A^2}{\mu_{A^2}}\right],\tag{10}$$

where $\Gamma(\cdot)$ denotes the Gamma function. The estimation of the mean power μ_{A^2} can be done with the MLE approach as explained previously for the Rayleigh distribution, while the shape parameter β_N can be estimated by using the estimator presented in [47], i.e.,

$$\tilde{\beta}_{N} = \begin{cases} \frac{(0.5000876 + 0.1648852y - 0.0544274y^{2})}{2}, \\ \text{if } 0 < y < 0.5772 \\ \frac{8.98919 + 9.059950y + 0.9775373y^{2}}{y(17.79728 + 11.968477y + y^{2})}, \\ \text{if } 0.5772 < y < 17 \end{cases}$$
(11)

where $y = \ln\left(\frac{\tilde{\mu}_{A^2}}{F}\right)$ and $F = \left(\prod_{i=1}^n A_i^2\right)^{\frac{1}{n}}$, and *n* is the number of samples considered in the estimation. The analytical equation of the *K* pdf is given by:

$$K_{pdf}(A) = \frac{4}{\Gamma(\mu_{A^2})} \left(\frac{\beta_K}{\mu_{A^2}}\right)^{\frac{\beta_K+1}{2}} A^{\beta_K} B_{\beta_K-1} \left[2A\sqrt{\frac{\beta_K}{\mu_{A^2}}}\right],$$
(12)

where B_{β_K-1} is the modified Bessel function of the second kind of order $\beta_K - 1$. The parameters of the K pdf can be estimated with the MLE by maximizing the logarithm of the likelihood function $l_n(\beta_K, \mu_{A^2}; \forall A_i, i \in [1..n])$ of the K distribution [34], i.e.,

$$(\tilde{\beta}_K, \tilde{\mu}_{A^2}) = \operatorname*{arg\,max}_{\left(\beta_K, \mu_{A^2}\right)} \ln [l_n],$$
(13)

where $\ln(\cdot)$ is the natural logarithm function.

- The analytical equation of the *Gamma* pdf is given by:

$$G_{pdf}(A) = \left(\frac{x}{\alpha_G}\right)^{\beta_G - 1} \cdot \frac{e^{-\frac{x}{\alpha_G}}}{\alpha_G \Gamma(\beta_G)}.$$
(14)

The values of the scale $\tilde{\alpha}_G$ and shape $\tilde{\beta}_G$ parameters of the Gamma distribution can be estimated using the MLE as solutions of the simultaneous equations [33]:

$$\begin{cases} \hat{\alpha}_G = \frac{\bar{A}}{\hat{\beta}_G}, \\ \log(\hat{\beta}_G) - \psi(\hat{\beta}_G) = \log\left[\bar{A}/(\prod_{i=1}^n A_i)^{1/n}\right], \end{cases}$$
(15)

where $\psi(\cdot)$ is the di-gamma function and $\overline{A} = E\{A\}$.

REFERENCES

- [1] V. Bogorodsky, C. Bentley, and P. Gudmandsen, D. H. D. Reidel Publishing Company, Ed., 1984.
- [2] K. Matsuoka, J. MacGregor, and F. Pattyn, "Predicting radar attenuation within the Antarctic ice sheet," *Earth and Planetary Science Letters*, vol. 359, pp. 173–183, 2012.
- [3] P. Gogineni, "CReSIS Radar Depth Sounder Data," http://data.cresis.ku.edu/, 2012.
- [4] C. Hernandez, V. Krozer, J. Vidkjaer, and J. Dall, "POLARIS: ESA's airborne ice sounding radar front-end design, performance assessment and first results," in *Microwave Symposium Digest*, 2009. MTT '09. IEEE MTT-S International, June 2009, pp. 393–396.
- [5] T. Ono, A. Kumamoto, Y. Yamaguchi, A. Yamaji, T. Kobayashi, Y. Kasahara, and H. Oya, "Instrumentation and observation target of the Lunar Radar Sounder (LRS) experiment on-board the SELENE spacecraft," *Earth, Planets, Space*, vol. 60, p. 321332, April 2008.
- [6] R. Seu, R. J. Phillips, D. Biccari, R. Orosei, M. A., G. Picardi, A. Safaeinili, B. A. Campbell, J. J. Plaut, L. Marinangeli, S. E. Smrekar, and D. C. Nunes, "SHARAD sounding radar on the Mars Reconnaissance Orbiter," *J. Geophys. Res.*, vol. 112, May 2007.
- [7] CReSIS, https://data.cresis.ku.edu/data/rds/2010_Antarctica_DC8/CSARP_mvdr/20101104_06/, 2011.
- [8] _____, https://data.cresis.ku.edu/data/rds/2011_Antarctica_TO/CSARP_mvdr/, 2013.
- [9] C. Gifford, G. Finyom, M. Jefferson, M. Reid, E. Akers, and A. Agah, "Automated polar ice thickness estimation from radar imagery," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2456–2469, 2010.
- [10] N. Karlsson, D. Dahl-Jensen, P. Gogineni, and J. Paden, "Tracing the depth of the Holocene ice in North Greenland from radio-echo sounding data," Annals of Glaciology, vol. 54, no. 64, pp. 44–50, 2013.
- [11] A.-M. Ilisei and L. Bruzzone, "A model-based technique for the automatic detection of earth continental ice subsurface targets in radar sounder data," *IEEE Geoscience and Remote Sensing Letters (GRSL)*, vol. 11, no. 11, pp. 1911–1915, Nov 2014.
- [12] O. Eisen, "Inference of velocity pattern from isochronous layers in firn, using an inverse method," *Journal of Glaciology*, vol. 54, no. 187, pp. 613–630, 2008.
- [13] E. Waddington, T. Neumann, M. Koutnik, H. Marshall, and D. Morse, "Inference of accumulation-rate patterns from deep layers in glaciers and ice sheets," *Journal of Glaciology*, vol. 53, no. 183, pp. 694–712, 2007.
- [14] D. Crandall, G. Fox, and J. Paden, "Layer-finding in radar echograms using probabilistic graphical models," in 21st International Conference on Pattern Recognition (ICPR), 2012, pp. 1530–1533.
- [15] L. Sime, R. Hindmarsh, and H. Corr, "Instruments and Methods Automated processing to derive dip angles of englacial radar reflectors in ice sheets," *Journal of Glaciology*, vol. 57, no. 202, pp. 260–266(7), 2011.
- [16] J. Mitchell, D. Crandall, G. Fox, and J. Paden, "A semi-automatic approach for estimating near surface internal layers from snow radar imagery," in International Geoscience and Remote Sensing Symposium (IGARSS), 2013.
- [17] C. Gifford and A. Agah, "Subglacial water presence classification from polar radar data," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 4, pp. 853 868, 2012.
- [18] R. Drews, "Layer disturbances and the radio-echo free zone in ice sheets," The Cryosphere, vol. 3, no. 2, pp. 195–203, 2009.
- [19] G. Freeman, A. Bovik, and J. Holt, "Automated detection of near surface Martian ice layers in orbital radar data," in *IEEE Southwest Symposium on Image Analysis Interpretation (SSIAI)*, May 2010, pp. 117–120.
- [20] A. Ferro and L. Bruzzone, "Automatic extraction and analysis of ice layering in radar sounder data," IEEE Transactions on Geoscience and Remote Sensing (TGRS), vol. 51, no. 3, pp. 1622–1634, 2013.
- [21] —, "Analysis of radar sounder signals for the automatic detection and characterization of subsurface features," *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, vol. 50, no. 11, pp. 4333–4348, November 2012.
- [22] S. Fujita, H. Maeno, S. Uratsuka, T. Furukawa, S. Mae, Y. Fujii, and O. Watanabe, "Nature of radio echo layering in the antarctic ice sheet detected by a two-frequency experiment," J. Geophys. Res., 104(B6), pp. 13013–13024, 1999.
- [23] D. Millar, "Radio-echo layering in polar ice sheets and past volcanic activity," Nature, no. 5822, pp. 441-443, 1981.
- [24] E. Rignot, J. Mouginot, and B. Scheuchl, "Ice flow of the antarctic ice sheet," Science, vol. 333, no. 6048, pp. 1427–1430, 2011.
- [25] G. D. Robin, S. Evans, and J. T. Bailey, "Interpretation of radio echo sounding in polar ice sheets," *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 265, no. 1166, pp. 437–505, 1969.
- [26] P. Nienow and B. Hubbard, Surface and Englacial Drainage of Glaciers and Ice Sheets. John Wiley and Sons, 2005, vol. 4, pp. 2575–2586.
- [27] G. d. Q. Robin and D. Millar, "Flow of ice sheets in the vicinity of subglacial peaks," Annals of GLacioLogy, vol. 3, pp. 290–294, 1982.

- [28] K. Matsuoka, T. Furukawa, S. Fujita, H. Maeno, S. Uratsuka, R. Naruse, and O. Watanabe, "Crystal orientation fabrics within the Antarctic ice sheet revealed by a multipolarization plane and dual-frequency radar survey," J. Geophys. Res., vol. 108, no. B10, 2003.
- [29] R. Bell, F. Ferraccioli, T. Creyts, D. Braaten, H. Corr, I. Das, D. Damaske, N. Frearson, T. Jordan, K. Rose *et al.*, "Widespread persistent thickening of the East Antarctic Ice Sheet by freezing from the base," *Science*, vol. 331, no. 6024, pp. 1592–1595, 2011.
- [30] M. Siegert, "Antarctic subglacial lakes," Earth-Science Reviews, vol. 50, pp. 29-50, 2000.
- [31] D. Daniels, Ground Penetrating Radar, 2nd Edition, ser. Ground Penetrating Radar. Institution of Engineering and Technology, 2004, vol. 1.
- [32] C. Oliver and S. Quegan, "Understanding Synthetic Aperture Radar Images," SciTech Publishing, 2004.
- [33] C. Forbes, M. Evans, N. Hastings, and B. Peacock, Multivariate Distributions. John Wiley and Sons, Inc., 2010.
- [34] J. Lin, "Divergence measures based on the Shannon entropy," IEEE Transactions on Information Theory, vol. 37, no. 1, pp. 145 151, January 1991.
- [35] S. Moysey, R. Knight, and H. Jol, "Texture-based classification of ground-penetrating radar images," Geophysics, vol. 71, no. 6, pp. K111–K118, 2006.
- [36] T. Algra, "Compression of raw SAR data using entropy-constrained quantization," in *Geoscience and Remote Sensing Symposium*, 2000. Proceedings. IGARSS 2000. IEEE 2000 International, vol. 6, 2000, pp. 2660–2662.
- [37] C. Cortes and V. Vapnik, "Support-vector networks," in Machine Learning, 1995, pp. 273-297.
- [38] G. Camps-Valls and L. Bruzzone, Kernel Methods for Remote Sensing Data Analysis. John Wiley and Sons, Inc, 2009.
- [39] L. Bottou, C. Cortes, J. Denker, H. Drucker, I. Guyon, L. Jackel, Y. LeCun, U. Muller, E. Sackinger, P. Simard, and V. Vapnik, "Comparison of classifier methods: a case study in handwritten digit recognition," in *Conference on Pattern Recognition, 1994. Vol. 2 - Conference B: Computer Vision amp; Image Processing, Proceedings of the 12th IAPR International*, vol. 2, 1994, pp. 77–82.
- [40] U.-G. Kressel, "Advances in kernel methods," B. Scholkopf, C. J. C. Burges, and A. J. Smola, Eds. Cambridge, MA, USA: MIT Press, 1999, ch. Pairwise Classification and Support Vector Machines, pp. 255–268.
- [41] A. Ben-Hur and J. Weston, "A users guide to support vector machines," Tech. Rep., 2012.
- [42] CReSIS, ftp://ftp.cresis.ku.edu/data/rds/rds_readme.pdf, 2013.
- [43] F. Rodriguez-Morales, S. Gogineni, C. Leuschen, J. Paden, J. Li, C. Lewis, B. Panzer, D. Gomez-Garcia Alvestegui, A. Patel, K. Byers, R. Crowe, K. Player, R. Hale, E. Arnold, L. Smith, C. Gifford, D. Braaten, and C. Panton, "Advanced multifrequency radar instrumentation for polar research," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 5, pp. 2824–2842, May 2014.
- [44] J. Li, J. Paden, C. Leuschen, F. Rodriguez-Morales, R. Hale, E. Arnold, R. Crowe, D. Gomez-Garcia, and P. Gogineni, "High-altitude radar measurements of ice thickness over the Antarctic and Greenland ice sheets as a part of Operation IceBridge," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 2, pp. 742–754, 2013.
- [45] R. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote sensing of environment*, vol. 37, no. 1, pp. 35–46, 1991.
- [46] I. Tsang, J. Kwok, P. Cheung, and N. Cristianini, "Core vector machines: Fast SVM training on very large data sets." *Journal of Machine Learning Research*, vol. 6, no. 4, 2005.
- [47] J. Greenwood and D. Durand, "Aids for fitting the gamma distribution by maximum likelihood," Technometrics, 1960.